Accepted Manuscript

Frictional diversification costs: Evidence from a panel of fund of hedge fund holdings

Juha Joenväärä, Bernd Scherer

PII: S0927-5398(19)30017-9
DOI: https://doi.org/10.1016/j.jempfin.2019.01.011
Reference: EMPFIN 1096

To appear in: *Journal of Empirical Finance*

Received date : 19 June 2016
Revised date : 16 March 2018
Accepted date : 29 January 2019

Please cite this article as: J. Joenväärä and B. Scherer, Frictional diversification costs: Evidence from a panel of fund of hedge fund holdings. *Journal of Empirical Finance* (2019), https://doi.org/10.1016/j.jempfin.2019.01.011

***Separate Title Page**

Frictional Diversification Costs: Evidence from a Panel of Fund of Hedge Fund Holdings

Juha Joenväärä and Bernd Scherer[♦]

This Version: 13 March, 2018

ABSTRACT

We analyze the diversification choices of fund of funds (FoF). Diversification is not a free lunch – not available for every FoF. Instead we find a positive log-linear relation between the number of constituent funds in a fund of hedge fund (n) and the respective assets under management, (AuM). More precisely it takes the form: n^2 AuM. This relation is consistent with the predictions from a model of naïve diversification with frictional diversification costs such as due diligence costs. Finally, we demonstrate that individual FoFs diversifying more in line with our model's predictions deliver superior performance and fail less likely.

JEL Classification: F14, G14

♦ We would like to thank the seminar participants at the 2014 Chicago Quantitative Alliance Fall Conference, the 2013 Austrian Working Group on Banking and Finance and Oxford University as well as an anonymous referee for helpful comments. We are grateful to Mikko Kauppila and Mikko Perttunen for helping us with the data processing. We are grateful for the support of the Academy of Finland and OP Group Research Foundation. Juha Joenväärä, PhD, is an Assistant Professor at the University of Oulu and a Visiting Researcher at Imperial College, London. Juha.joenvaara@oulu.fi. Bernd Scherer, PhD is Chief Investment Officer at a managed futures hedge fund in Vienna and a Visiting Professor at WU Wien. drberndscherer@gmax.net.

*Highlights (for review)

Highlights

- Paper employs a model of naïve diversification with frictional diversification
- Provides evidence supporting model’s predictions

Frictional Diversification Costs:
Evidence from a Panel of Fund of Hedge Fund Holdings

This Version: 13 March, 2018

ABSTRACT

We analyze the diversification choices of fund of funds (FoF). Diversification is not a free lunch – not available for every FoF. Instead we find a positive log-linear relation between the number of constituent funds in a fund of hedge fund (n) and the respective assets under management, (AuM). More precisely it takes the form: $n^2 \propto AuM$. This relation is consistent with the predictions from a model of naïve diversification with frictional diversification costs such as due diligence costs. Finally, we demonstrate that individual FoFs diversifying more in line with our model's predictions deliver superior performance and fail less likely.

JEL Classification: F14, G14

1. Introduction

This study analyzes the diversification choices of fund of funds (FoF). FoFs arguably involve the most sophisticated subset of institutional investors. For this purpose, we extend the setup of Goldsmith (1976) to present a parsimonious log-linear, fixed-effects panel model for the optimal diversification of FoFs under frictional diversification costs. The optimal number of individual fund holdings in any FoF depends on the sum of assets under management (*AuM*), the costs of due diligence and/or monitoring (i.e., frictional diversification costs), and the respective (clientele) risk aversion. Brown et al. (2008a) describe due diligence as an expensive activity. Consequently, larger funds can more easily absorb that cost. Attitudes toward risk also impact the diversification choice: risk-averse investors are willing to incur more frictional diversification costs for a small reduction in risk than less for risk-averse investors.

We test our empirical model on a unique set of 10 years of quarterly data for a cross section of 127 FoF's. These panel data are obtained directly from SEC filings and are not available from commercial databases (e.g., Lipper TASS or BarclayHedge). We find strong evidence that frictional diversification costs limit the applicability of traditional diversification advice.

In our empirical analysis, we use a set of fixed-effects panel regressions. In line with our theoretical predictions, increasing levels of assets under management allow FoFs to increase their number of holdings with an elasticity of 0.5. This is consistent with a world where frictional diversification costs will force smaller FoFs to offer a less diversified portfolio of hedge funds than their larger peers would. In the absence of those costs, we might still find different levels of diversification related to fund manager conviction or predictive capabilities but not related to the fund size.

Our model survives a battery of robustness tests. First, we explore the robustness of fixed-effects panel regressions by dropping out potentially influential data points or individual FoFs. Second, we compare the fixed-effects regression with a pooled ordinary least-squares (OLS) regression. Without individual effects, the data would imply that FoFs do not differ with respect to frictional costs, risk aversion, or investment skill.[1] However, we can reject the pooled OLS model that ignores individual effects which seem to be important. Next, we turn to nonlinearities or threshold effects which might be important. It could be the case that when assets under management grow, a FoF can increase the number of its holdings because frictional diversification costs become proportionally smaller. Therefore, the predicted theoretical

[1] Fung, Hsieh, Naik and Ramadorai (2008) show that a subset of FoFs poses investment skill (i.e., delivers a significant Fung and Hsieh (2004) alpha).

relation may not be log-linear. We find no empirical evidence for this either. To sum up, we provide robust evidence that FoFs tend to set portfolios with $n^2 \propto AuM$.

We next relate the diversification to the FoF future performance. Our conjecture is that over-diversification (too high costs, i.e., too low expected return per unit of risk or sloppy due diligence if costs are not spend) is as detrimental to performance as under-diversification (too high risk per unit of expected return). Using both univariate and multivariate analyses, we document a robust statistical relationship between the degree of diversification and FoF future performance. Indeed, the FoFs diversifying more in line with our model's predictions deliver superior performance. The economic significance of this finding is large and it cannot be subsumed by other variables that has been documented to explain FoF performance. According to our multivariate regression results, FoFs that do not diversify along simple model deliver 1.40%–3.40% lower returns per annum compared to the FoFs that display diversification consistent with our model of diversification under frictional costs. Hence, even after controlling for the FoF's size (Brown, Frazier, and Liang 2008), the number of underlying funds (Brown, Gregoriou, and Pascalau 2012), the FoFs' capital flows (Fung, Hsieh, Naik and Ramadorai 2008), past performance or autocorrelation in returns (Getmansky, Lo, and Makarov 2004), a proxies for operational risk (Brown, Goetzmann, Liang, and Schwarz 2008b, 2009, 2012), compensation structure (Agarwal, Daniel, and Naik 2009), and discretionary (Aiken et al 2015b) and non-discretionary liquidity restrictions (Aragon 2007), we find that our performance results holds suggesting the FoFs diversifying closely along the lines of our simple model deliver superior future performance.

Finally, we examine the relationship between the degree of diversification and operational risk.[2] Our results reveal that individual FoFs diversifying more in line with our model's predictions exhibit a lower failure probability than their over- or under-diversified peers even after controlling for the role of existing operational risk measures. To better understand mechanism behind the failure of FoFs, we examine the link between operational risk and FoFs' underlying holdings. We find the FoFs diversifying more in line with our model's predictions hold less often individual funds that have imposed discretionary liquidity restrictions such as side pockets and gates. We, however, are not able to document a significant relationship between the degree of diversification and operational problems (Brown, Goetzmann, Liang, and Schwarz 2008b) or suspicious patterns in fund returns (Bollen and Pool 2012).

Our work is related on several steams of existing literature. Previous work on FoFs is devoted mainly to one simple question: How many hedge funds are needed for a *diversified* fund of funds or a hedge fund

[2] We are grateful for referee about this suggestion.

portfolio? Henker and Martin (1998), Amin and Kat (2002), Lhabitant and Learned (2002), and Brown et al. (2012) all use a simple two-step procedure to test for over- or underdiversification.

> Step 1: Simulate random portfolios of increasing size (i.e., increasing number of equally weighted assets) and plot the evolution of volatility as a "diversification curve": a functional relationship between portfolio standard deviation and portfolio size.[3]
>
> Step 2: Decide when the marginal improvement in the statistic derived during step 1 becomes "small".

What "small" means is usually determined by eyeballing the diversification curve, so that it tends to reflect the researcher's subjective judgment (or perhaps his eyesight). These papers typically find that the optimal number of hedge funds ranges between 5 and 25. Without an explicit model from which to argue how much diversification is warranted in the presence of diversification costs this looks like an ambitious statement. We conclude that methods based on diversification curves have three major shortcomings. First, no attempt is made to specify the frictional costs of adding another fund to a portfolio. In the absence of such costs it is always optimal to naively diversify across all possible investments. In the early literature, Samuelson (1967) and Brennan (1975) make this point by stating that investors should diversify as much as possible while remaining aware of the trade-off between diversification and its costs. However, there is no study that formally incorporates the frictional diversification costs faced by real-world investors. We fill this research gap. Frictional costs comprise the costs, per each additional fund, of due diligence and monitoring as well as the loss of the power to bargain for fee rebates when diversifying among too many funds. The second main deficiency in the "diversification curve" method is that it does not account for the actual assets under management – even though fixed costs can be spread more easily across a large pool of assets. A decision maker seeking the optimal number of assets in which to invest $10 million versus $100 million should certainly receive a different answer in each case. Third, the reduction in volatility that diversification is intended to provide is most valuable for investors with high risk aversion. It is clear that investors with low risk aversion will be less willing to pay the diversification costs of reducing risk (i.e., reducing volatility by adding more funds).

[3] In the case of volatility, Elton and Gruber (1977) show that a closed-form solution exists. Yet the simulation aspect is useful for illuminating how this procedure might extend to measures of risk and performance that are more complicated.

Our research differs from that of authors who look to explain poorly diversified portfolios (portfolios with too little a number of names) in terms of the behavioral shortcomings of private versus professional decision makers. Statman (2004) coined the term "behavioral portfolio theory": the attempt to (psychologically) rationalize the observed underdiversification of individual investors. In Statman's view, individual investors divide their total wealth into mental "buckets" according to their investment goals. Equities fall into the top portfolio layer, which reflects the investor's desire for returns to a "lottery ticket" scale. Recent support for this perspective is provided by Frazzini and Pedersen (2014), who find that aversion to leverage (a form of risk aversion) leads investors to construct underdiversified portfolios that concentrate on more volatile stocks. Elton et al. (2004), Plokovnichenko (2005), Mitton and Vornick (2007), and Phillips et al. (2007) all expand this theme of positive skewness and lottery preferences. Of course, a lack of diversification might simply reflect frictional diversification costs. We would then expect large funds (as measured by *AuM*) to display more diversification. This is the previously unexamined focus of our study.

The work reported here is related to the growing set of papers addressing hedge fund operational risk and role as financial intermediaries. In a series of papers, Brown, Goetzmann, Liang, and Schwarz (2008b, 2009, 2012) show that hedge funds with a higher operational risk tend to deliver lower average performance and to exhibit a greater likelihood of failure. Aiken et al. (2015a) show that FoFs may provide valuable due diligence and monitoring services for investors by firing underperforming managers. Agarwal, Nanda, and Ray (2013) examine hedge fund investments of institutional investors. They find that larger institutions enjoy economies of scale, enabling direct investment into relatively better performing hedge funds. Brown et al. (2012) establish that FoFs suffer from overdiversification and that this condition may well be associated with their inability to perform timely due diligence, which is costly when a FoF invests in a large number of hedge funds. We differ from that literature by explicitly modelling the FoFs' portfolio choice with frictional costs related to due diligence costs.

The rest of this paper is organized as follows. Section 2 derives a parsimonious empirical model for the optimal number of holdings in the presence of frictional diversification costs. Section 3 describes our data set, after which Section 4 presents the empirical tests. Section 5 examines diversification and fund performance as well as operational risk. Section 6 concludes.

2. A Parsimonious Model of FoF Diversification

Following Goldsmith (1976) and Scherer (2013), we assume that a FoF employs a naïve decision maker who has no information on returns or risks. This decision maker will trade off the fund's marginal benefits from naïve diversification (formulated as marginal risk reduction multiplied by risk aversion) against their marginal costs to diversify, which we view as the frictional costs that arise from due diligence and monitoring. Diversification is measured by the number of hedge funds in a FoF. Despite this obviously heuristic way of conceiving diversification, Goetzmann and Kumar (2008) show that investors succeed (i.e., their returns increase) under naïve diversification but not under optimized diversification – that is, when constructing a portfolio based on assets' volatility and correlation.[4] We model optimal diversification for a standard mean variance investor. In our case this investor looks for a solution of

(1)
$$n^{*} = \arg\max_{n} \mu - \lambda\sigma^{2}(n) - n\frac{f}{\mathrm{AuM}}$$

Here $\sigma^2(n)$ denotes the risk (variance) of an equally weighted portfolio of size n, λ the investor's risk aversion, and $\frac{f}{\mathrm{AuM}}$ the additional costs (i.e., fixed costs f per additional fund as a fraction of assets under management, AuM). We can most simply think of f as the costs of a due diligence report. The costs of exercising due diligence are far from trivial, with industry insiders estimating them to range between \$50,000 and \$100,000 (US).[5] However, this would make it a one off expense which does not sit well with our panel data regression. The latter explicitly estimates within-funds effects, i.e., the reaction of a FoF to an increase or decrease of its assets under management. Instead we could interpret f as per time interval costs of holding a hedge fund either in the form of due diligence costs spread across the expected hedge fund holding period or as monitoring and complexity costs.

A few additional remarks on the implicit assumptions of our model are in order. First, our investor is unable to form differential estimates on either returns, volatility or correlation. He replaces individual estimates with universe wide averages on returns, correlations and volatilities.[6] While 1/n investors are typically investors that have no information at all, our investor must at least have some information on

[4] In our context assets' mean return, volatility and correlations are based on individual hedge fund returns.

[5] See Greenwich Associates (2011). To the extent that due diligence costs are expected to fall (because of a secondhand market for due diligence reports), diversification will increase. According to Brown et al. (2012), due diligence reports are cheaper when investors are willing to share them.

[6] Although the ex post spreads might be large between funds, the ex ante predictability is low (Bollen, Joenväärä, and Kauppila 2018) and difficult to exploit due to share restrictions (Joenväärä, Kosowski and Tolonen 2018). However, the cross-correlations may not be similar between funds. Fortunately, several papers suggest that the accurate estimates for expected returns and variances are much more important than correlations (e.g. Chopra and Ziemba 1991). We therefore believe that our assumption are realistic.

universe averages. Otherwise he could not trade of the marginal increase in utility (from a marginal reduction in portfolio risk) with the marginal increase in diversification costs. Second, our model is a one period model. While this supports the interpretation of f as due diligence costs, we can view f more broadly as coordinating or monitoring costs that also apply in a multi-period context.

Another consequence from the one period character of the model is that we will exclude the fund managers incentives (maximize his fee income) from the model by assuming the manager of the FoF always acts in the clients' best interest. Without this assumption performance fee maximization would lead the FoF manager to take excessive risk with no interest in diversification. Once we introduce performance based on fees, the portfolio management always has an incentive to increase risks in a one period model (with or without frictional diversification costs). Due diligence costs would make diversification even less attractive as these costs will work as a drag on performance and hence reduce expected performance fees. In contrast to this one period insight, the theoretical and empirical literature has shown that once we move to a multi-period setting – that allows us to introduce performance related job losses or outflows after performance blowouts- the one period intuition to increase risks is largely mitigated. Fung and Hsieh (1997) and Brown, Goetzmann and Park (2001) find empirical evidence, that reputational concerns largely dampen risk taking incentives from one period models. Theoretical work by Xu and Scherer (2007) and Panageas and Westerfield (2009) confirms the empirical evidence. A one period intuition does not carry over to (real world like) multi-period environments.

Even if the existence of incentive fees would have a material effect on risk taking in the real world, we believe it is not relevant for two reasons. First, in our panel data regression individual effect will take care of the individual differences between FoFs. Second, we would not expect this to affect the relationship between the number of funds and assets under management, unless the incentive fee design covaries with the fund size.

We can now write down the first-order condition of our investor above as

(2) $$\frac{f}{\text{AuM}} = -\lambda \frac{d\sigma^2(n)}{dn};$$

The expected variance for an equally weighted portfolio is well known to be[7]

[7] Elton and Gruber (1977) prove that that this equality holds as an expectation if funds are selected randomly (i.e., without prior knowledge).

(3) $$\sigma^2(n) = \frac{\overline{\sigma}^2}{n} + \left(1 - \frac{1}{n}\right)\overline{\sigma}^2\overline{\rho},$$

where $\overline{\sigma}$ and $\overline{\rho}$ are the average volatility and correlation in the universe of investable assets. We can find an explicit solution for the marginal change in risk,

(4) $$\frac{d\sigma^2(n)}{dn} = \frac{1}{n^2}\overline{\sigma}^2(\overline{\rho} - 1).$$

Substituting (4) into (2) yields $\frac{f}{\mathrm{AuM}} = -\lambda\frac{1}{n^2}\overline{\sigma}^2\left(\overline{\rho} - 1\right)$, which can be solved for the optimal n:[8]

(5) $$n^* = \sqrt{\lambda\overline{\sigma}^2\left(1 - \overline{\rho}\right)\left(\frac{f}{\mathrm{AuM}}\right)^{-1}}.$$

The optimal number of assets increases with rising risk aversion (λ), rising average volatility ($\overline{\sigma}^2$), falling average correlation ($\overline{\rho}$), falling frictional costs (f), and rising *value* of assets under management (AuM). A portfolio with a small *number* of assets need not be under-diversified. It could simply be a small portfolio (low assets under management), or it might belong to investors who are less averse to risk or who would incur high due diligence costs per additional fund.

For our naïve investor facing frictional diversification costs, the risk of an optimally diversified portfolio (i.e., one for which the marginal benefits from diversification only just equal the costs of diversifying) is found by substituting (5) into (3):

(6) $$\sigma^2(n^*) = \overline{\sigma}^2\overline{\rho} + \sqrt{\frac{\overline{\sigma}^2(1-\overline{\rho})(f/\mathrm{AuM})}{\lambda}}.$$

The first term on the right-hand side of (6) represents the average covariance risk in the available asset universe. This is the minimal achievable risk for $n \to \infty$ – that is, in the absence of frictional costs or for investors who are infinitely averse to risk (rendering frictional costs unimportant). Note, that $n \to \infty$ is in general not the optimal strategy for optimized diversification. The number of holdings in a minimum variance portfolio with known sample covariance under a long only constrained will not generally expand with the size of the universe. The second term reflects the higher risk that results when a diversification strategy accounts for its associated frictional costs, since those costs preclude investors from diversifying to the theoretical maximum. Taking logs on both sides of (5) results in a linear model,

[8] This expression is identical to equation (1.10) in Goldsmith (1976, p. 1130). Despite its convincing intuition, that model was not adopted by the empirical literature and has been largely ignored in both academic and practical work. The rest of this section is devoted to extending the model's conclusions and shaping it into a testable form.

(7) $$\log(n) = a + b \cdot \log(\text{AuM}) + \varepsilon\,,$$

where $a = 0.5 \cdot \log\left(\lambda\bar{\sigma}^2(1-\bar{\rho})/f\right)$ and $b = 0.5$.[9] However, our proposed model predicts that first, there is a positively sloped relationship between the number of funds and the amount of assets under management. Second, it predicts that this relation is not linear but log-linear, with a slope coefficient of 0.5. In a log-linear model, $\hat{b}$ signifies elasticity (here, the percentage increase in number of assets for each percentage increase in assets under management). In order to test (7) on our unique panel data set, we propose running a fixed-effects model of the form

(8) $$\log\left(n_{it}\right) = a_i + b \cdot \log\left(\text{AuM}_{it}\right) + \varepsilon_{it}\,,$$

where i denotes a specific FoF and t a particular moment in time. This choice is motivated by the possibility of omitted variable bias in (7). Equation (7) is not likely to hold when investors are strongly convinced of their own forecasting ability; such conviction is the natural enemy of diversification. The better our forecasting abilities, the more concentrated (i.e., less diversified) our optimal portfolios will become. Hence it is safe to assume that investors with actual or even presumed forecasting skills will hold fewer funds than do investors with weaker forecasting abilities. Greater investment skills manifest as variation in individual effects (a_i). Investors who are optimistic about their level of forecasting ability will invest in a fewer hedge funds irrespective of the amount of assets under management. So even though (7) suffers from that misspecification, our fixed-effects, panel data model uses cross-sectional units as controls. The consequences of unobserved investment skill should cancel out provided the effect of skill is constant (i.e., a fixed effect). Empirical facts also support our specification, because Fung, Hsieh, Naik and Ramadorai (2008) document that a subset of FoFs consistently delivers alpha or poses investment skill. Another neat side effect of our panel data model is that it corrects for alternative investment universes, clientele effects (risk aversion) and frictional diversification costs.

3. Data

To test the empirical predictions generated by our model of naïve diversification under frictional costs, we use a panel of registered fund of (hedge) fund holdings for the period 2003Q1–2012Q4. A FoF may opt to

---

[9] Instead of using $n\frac{f}{AuM}$ we could also use $n^k\frac{f}{AuM}$ more generally for modelling total frictional costs. This would lead to $a = \frac{1}{k+1}\log\left(\frac{\lambda_{it}\bar{\sigma}^2(1-\bar{\rho})k}{f}\right)$ and $b = \frac{1}{k+1}$. If costs functions differ, slopes across FoFs would also differ. For $k < 1$ (economies of scale in information gathering) $b > \frac{1}{2}$ and hence the number of funds rises faster with increasing assets under management. This would look like overdiversification, but it is not.

register with the US Securities and Exchange Commission (SEC) under the Investment Company Act of 1940, thereby gaining wider distribution channels. Registered FoFs are considered to be closed-end funds and so are usually not listed on exchanges. Exactly as do mutual funds, registered FoFs must disclose mandated filings publicly, including quarterly disclosures of portfolio holdings and semi-annual financial statements. Following Aiken et al. (2013, 2015a, 2015b), we gather the underlying hedge fund holdings of our sample FoFs from SEC forms N-Q, N-CSRS, and N-CRS. The data in these filings enables us to create a panel of quarterly hedge fund holdings. For each FoF, the panel contains the current value of each position. Hence we can calculate each FoF's total assets under management (AuM) and number of hedge funds (n) on a quarterly basis. From these filings, we are also able to gather information on underlying hedge funds' both discretionary and non-discretionary liquidity restrictions as well as on their investment strategy. Next, we gather N-2 filings for each of the FoF's in our sample. From N-2 filings, we hand-collect manually information on compensation structure and share restrictions at the FoF-level. Finally, we merge the FoFs' underlying funds to union of commercial databases (BarclayHedge, EurekaHedge, HFR, Lipper TASS and Morningstar) as well as to Form ADVs. This allows us to gather data on individual funds' quarterly returns, assets, fund-specific characteristics as well as funds' operational problems. Appendix A provides more details about data gathering and merging process.

It is important to emphasize that commercial hedge fund databases do not provide panel data about the number of underlying hedge funds in which a FoF has invested. Therefore, it would be impossible to test our model predictions of frictional diversification costs if we were limited to using data obtained from commercial databases. Some commercial databases (e.g., BarclayHedge, EurekaHedge) do provide a "snapshot" view of how many individual funds are in a given FoF's portfolio, but none of them provide information on how that number of funds changes over time. For example, the AuM of FoFs increased rapidly *before* the financial crisis but redemptions were rampant *during* that crisis. Finally, because our data contains actual investors' holdings that are investable, it exhibits neither survivorship nor backfilling bias – both of which are typical of commercial databases (see, e.g., Fung and Hsieh (2000), Liang (2000), Agarwal et al. 2013). However, a potential concern is that our sample of FoFs may introduce a different form of selection bias because only a subset of FoFs is registered with the SEC. Fortunately, Aiken et al. (2013) show that the returns of registered FoFs do not seem to differ from those of the FoFs that report to the commercial databases (BarclayHedge, HFR and Lipper TASS). Indeed, registered FoFs are often run by the most prominent hedge fund management firms that are rarely available for researchers (Edelman, Fung and Hsieh 2013).[10] We therefore believe that our novel data and model provide a fertile setting in

[10] Agarwal, Lu and Ray (2016) find that only 8 of registered FoFs report voluntary to the Lipper TASS database.

which to explore the possibility of sophisticated institutional investors constructing portfolios that provide optimal diversification benefits even after frictional costs are taken into account.

Our raw data are summarized in Table 1 and represented graphically in Figure 1 and
Figure 2. Both the dependent and the independent variables exhibit considerable variation across and also within units (FoFs). This feature of the data is important because it enables using a fixed-effects panel regression to address a potential omitted variable bias. We sort the 127 FoFs into five groups depending on the amount of available data. Because there are few observations for many of our sample FoFs, we are using an unbalanced panel. Of these 127 FoFs, 49 have less than one year of data available (about 2 quarters on average), and 76 FoFs have less than two years of quarterly data. The table includes (as row 5) the individual regression slopes derived from (7). However, the small number of observations virtually guarantees that individual regressions will yield unreliable estimates.

We therefore assume that the regression slopes are similar but not identical and apply the mean group estimator instead. In each group, for every FoF we estimate $\log(n_i) = a_i + b_i \log(\mathrm{AuM}_i) + e_i$. Let T_i denote the number of available observations for the ith FoF; then the mean group estimator for all FoFs in group k is given by

(9) $$\hat{b}_k = \left(\sum_{i=1}^{N(k)} T_i \hat{b}_i\right)\left(\sum_{i=1}^{N(k)} T_i\right)^{-1},$$

where $N(k)$ is the number of FoFs in this group. The variance of (9) is calculated as follows:

(10) $$\mathrm{var}\left(\hat{b}_k\right) = \left(\sum_{i=1}^{N(k)} T_i^2 \,\mathrm{var}\left(\hat{b}_i\right)\right)\left(\sum_{i=1}^{N(k)} T_i^2\right)^{-1}.$$

The mean group estimator is based on weighting individual regression slopes by the number of available observations. The variance of the mean estimate depends not only on the number of observations but also on the precision of each estimate. Except for the FoFs with long observation histories, all estimates are noisy and few differ significantly from our conjecture of 0.5.[11]

[11] Performing robust regressions as suggested by Huber (1981) does not change our results.

Table 1

Data Characteristics

Funds of funds are sorted (by the number of available observations) into five sample size groups. For each group we report the number of FOFs it contains, the average number of observations, the average of $\log(\mathrm{AuM})$, and the average number of holdings. We also estimate the mean group estimator for the average regression slope in $\log(n_i) = a_i + b_i \log(\mathrm{AuM}_i) + e_i$ for all FoFs in each group. Given T_i observations (for FoF i), the mean group estimator for all FoFs in group k is given by $\hat{b}_k = \left(\sum_{i=1}^{N(k)} T_i \hat{b}_i\right)\left(\sum_{i=1}^{N(k)} T_i\right)^{-1}$ for $N(k)$ the number of FoFs in the group. The variance of the mean group estimator is given by $\mathrm{var}(\hat{b}_k) = \left(\sum_{i=1}^{N(k)} T_i^2 \mathrm{var}(\hat{b}_i)\right)\left(\sum_{i=1}^{N(k)} T_i^2\right)^{-1}$.

	$\#\mathrm{Obs.} < 4$	$4 \leq \#\mathrm{Obs.} < 8$	$8 \leq \#\mathrm{Obs.} < 12$	$12 \leq \#\mathrm{Obs.} < 20$	$20 < \#\mathrm{Obs.}$
Number of FoFs, $N(k)$	49	27	17	20	14
Average number of observations	1.86	5.67	9.[illegible]7	15.3	23.93
Average of $\log(\mathrm{AuM})$	17.63	18.23	18.71	18.1	18.63
Average number of holdings, n	18.91	25.72	26.5[illegible]	21.2	31.14
Mean group estimator (MGE) of $\hat{b}$	0.12	0.49	0.66	0.55	0.37
Standard Deviation of MGE, $\sigma(\hat{b})$	0.21	0.27	0.09	0.0[illegible]	0.01

Figure 1

Number of Holdings per FoF

The boxplots in this figure represent the distribution of the number of hedge fund holdings for each of our 127 FoFs. Entries are sorted from left to right by the number of available fund observations. Printed at the top of each grouping are the number of quarterly observations available for – and the number of FoFs belonging to – that group.

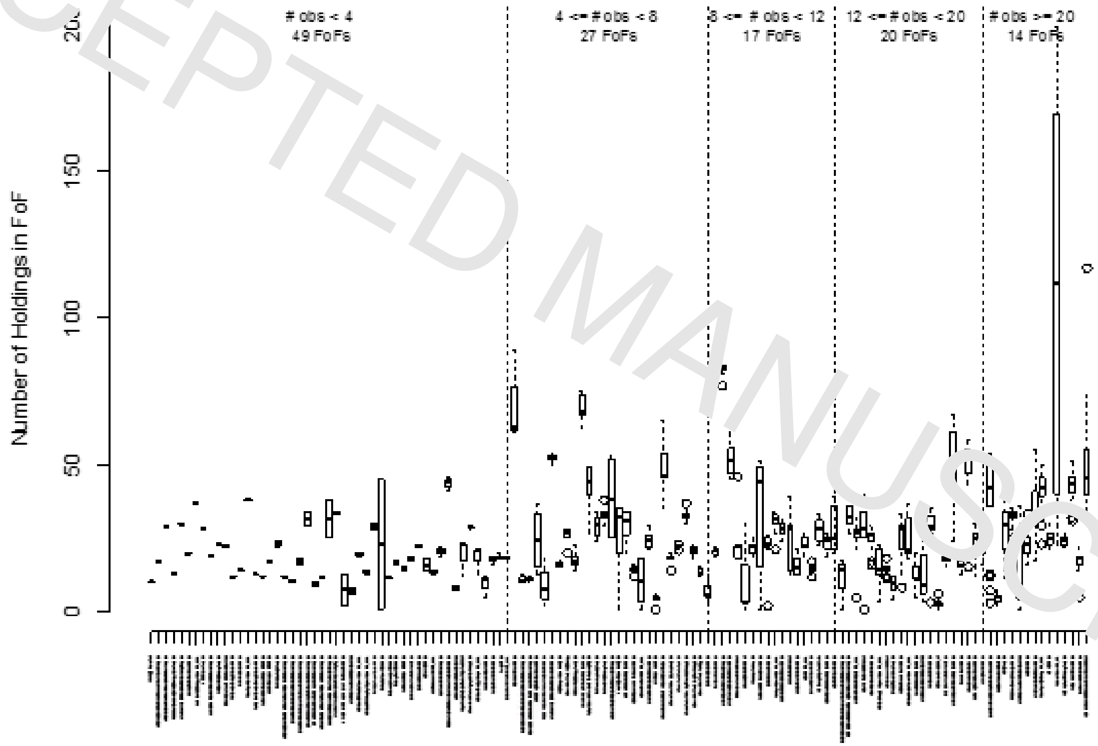
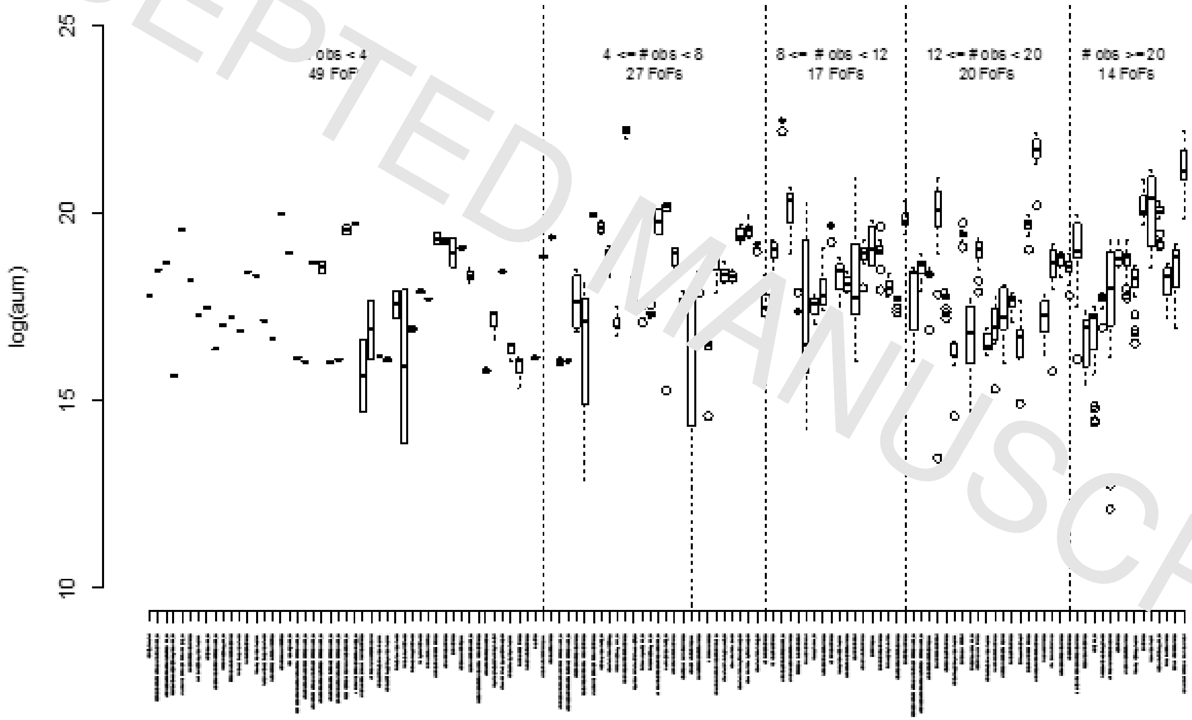


Figure 2

Assets under Management per FoF

The boxplots in this figure represent the distribution of $\log(\text{AuM})$ for each of our 127 FoFs. Entries are sorted from left to right by the number of available fund observations. Printed at the top of each grouping are the number of quarterly observations available for – and the number of FoFs belonging to – that group.



4. Empirical Tests

In this section, we test the predictions of the model. We first use a fixed-effects panel regression model to examine whether FoFs are constructed in line of our naïve predictions. We then explore alternative specifications as well as perform various robustness tests.

4.1. Baseline Fixed-Effects Panel Regressions

We start by running a fixed-effects (FE) panel regression model as suggested by our theoretical discussion in Section 3. Equation (11) presents our estimates for $i = 1,...,127$ and for time t ranging from 2003Q1 to 2012Q4, where the statistical *t*-value is given in brackets:[12]

(11)
$$\widehat{\log\left(n_{it}\right)} = a_i + \underset{[12.843]}{0.473}\log\left(\text{AuM}_{it}\right), \quad \bar{R}^2 = 0.85\,.$$

For this FE model the estimated slope is highly significant (*t*-value of 12.843) and, at 0.473, comes close to our prediction of 0.5. A Wald test for whether $\hat{b}$ is indeed insignificantly different from our prediction of 0.5 (with $H_0 : b = 0.5$) results in a $\chi^2(1)$-distributed test statistic of 0.54. The corresponding *p*-value of 0.45 does not allow us to reject the null hypothesis of $b = 0.5$. Including time effects does not change this result, although the estimated slope coefficient for the log of AuM does change marginally (to 0.503; *t*-value of 17.4). Yet testing against a slope of 0.5 results in a *p*-value of 0.97 – that is, making it even more difficult to reject our theoretical prediction. We also try omitting all funds with less than 1 or 2 year's worth of observations; although the slope estimate then drops to 0.45 or 042, it remains economically close to our conjectured value of 0.50. Indeed FoF managers seem to set $n^2 \propto AuM$.

One issue with our analysis might be the large number of funds (76 out of 127) with less than 8 observations. Typically these funds also display high persistence (small variations) in both dependent and independent variables. Low within-fund variation will boost model fit (as fund specific intercepts are sufficient to reduce the funds contribution to regression errors) and increase $\bar{R}^2$ in an unbalanced panel regression. At the same time the sampling error for the slope coefficient is considerably higher than for a standard pooled regression model. However this is

[12] Standard errors for significance tests in both regressions are corrected for cross-sectional heteroskedasticity by using a robust covariance matrix as in Greene (2008, p. 185). This correction was necessitated by a likelihood ratio test on the equality of variances across units (FoFs). That test rejected the null hypothesis of homogeneous variance with a *p*-value of 0.000, which is only philosophically different from zero.

unlikely the case if within fund variation is driven by a common rule across all FoF's. It is exactly the (artificially) small precision in the slope coefficient that might avoid the rejection of $H_0 : b = \frac{1}{2}$. The values for a_i fall between -6.21 and -4.74. Are they realistic? For example if $\sigma = 0.2, \rho = 0.5, f = 25000, \lambda = 5$ we arrive at $a = -6.21$.

How much does our estimate vary when individual units (FoFs) are dropped from the panel data set? We repeat the estimation of (11) for a total of 127 times each time dropping (a different) one of the units (FoFs) from the sample. Figure 3 plots the variation in estimates. As in previous figures, all entries are sorted from left to right by the number of available fund observations. As expected, omitting funds with few observations (first column in Table 1) has little or no effect on the estimated coefficients. Variation increases as FoFs with more observations are sequentially dropped from the sample. Yet all coefficient estimates remain near the conjectured value of $b = 0.5$, and none becomes statistically different from 0.5. Hence none of the FoF's solely drives our results.

Finally, we check for the existence of influential data points capable of driving the results. For each observation in our FE panel regression, we calculate its Cook's distance (sum of squared differences between full-sample fitted values and fitted values from a model leaving one observation out, standardized by the number of parameters *times* the model's mean squared error). This distance measure is plotted in Figure 4. To check for the collective effect of influence points we drop the 1% of observations with the highest distance measure from our sample and then repeat estimating (11).

(12) $$\widehat{\log(n_{it})} = a_i + \underset{[30.223]}{0.455} \log(\mathrm{AuM}_{it}), \quad \bar{R}^2 = 0.92 .$$

4.2. Alternative Models and Robustness Tests

We next investigate whether alternative regression model specifications or functional forms fit better in with the data than with the fixed effects regression model. In line of our theoretical discussion, we show that the fixed effects regression model provides better fit in the data than the pooled regressions or alternative functional forms.

In order to validate our conjecture about the presence of individual effects in FoFs, we compare the FE regression with a pooled ordinary least-squares (OLS) regression. The absence of individual effects would imply that FoFs do not differ with respect to frictional costs, risk aversion (a reflection of different clienteles), or investment skill. The fit to this pooled OLS regression is given by

$$\widehat{\log\left(n_{it}\right)} = \underset{[-12.448]}{-3.293} + \underset{[24.741]}{0.345}\log\left(\mathrm{AuM}_{it}\right), \quad \bar{R}^2 = 0.44. \tag{13}$$

The original data and the fitted values for both regressions are plotted in Figure 5. We perform an *F*-test of fixed effects versus pooling by comparing the residual sum of squares for both models. With a test statistic of 20.14, the *p*-value is close to zero; hence we can reject the pooled OLS model. Pooling all observations amounts to ignoring individual effects, which will bias the estimated slopes (0.473 for FE regression versus 0.345 for pooled OLS regression).[13] This suggests that we should not ignore individual effects regarding the FoF specific the frictional costs, the risk aversion, or the investment skill.

Although we have shown that the fixed-effects panel model is both theoretically and statistically the most proper way to test the predictions of our simple model, it is, however, interesting to examine the determinants of diversification choice. Table 2 present results for pooled OLS with a set of control variables. Once we add the control variables, the coefficient for $\hat{b}$ is 0.258 with a statistic of 26.74. More interestingly, one potential major determinant of diversification choice is the fee structure.[14] We, however, find that both the coefficient for management fee and the coefficient for incentive fee are only marginally significant, while the coefficient for sales load fee is statistically insignificant. Among the share restriction variables, only the coefficient for notice period is positive and significant, whereas the coefficients for lockup period and redemption period are insignificant. Finally, we find a negative relationship between the style concentration and diversification. However, this relationship is rather mechanical, since the FoFs with a fewer underlying funds are also concentrated.

[13] We also compare the fixed-effects model with a random-effects (RE) model by means of the Hausmann test. The $\chi^2(1)$-distributed statistic takes a value of 21.27, so we can reject the null hypothesis of a RE panel model with high confidence (*p*-value of 0.000). Our slope parameter for the RE model is 0.47, which is both numerically and statistically close to our conjectured value of 0.5. A Wald test with $H_0 : b = 0.5$ results in a $\chi^2(1)$ of 2.26 (i.e., a *p*-value of 0.129).

[14] We thank for anonymous referee about suggestion.

Table 2
Pooled OLS with Control Variables

This table present pooled OLS regression results, in which log(number of funds) is explained by log(AuM), compensation structure variables (management fee, incentive fee and sales load fee), share restrictions (lockup dummy, redemption period, notice period), and style concentration.

	Log(Number of funds (n))
log(AuM)	0.25[illegible]
	(26.74)
Management fee	0.80[illegible]
	(1.87)
Sales load fee	-0.300
	(-0.72)
Incentive fee dummy	-0.058
	(-1.72)
Lockup dummy	-0.002
	(-0.08)
Redemption period	-0.076
	(-1.03)
Notice period	1.040
	(5.10)
Payout period	-0.421
	(-0.80)
Style concentration	-1.339
	(-20.96)
Intercept	-1.103
	(-4.77)
Adjusted R-square	0.659
Number of Observations	914

Because our model suggests a linear regression in logs, we can use the MacKinnon et al. (1983) test to see whether a regression in logs does deliver better results than are obtained from a regression using untransformed data. In particular, we compare the log-linear model with a model using untransformed data, $n_{it} = \alpha_i + \beta \cdot \mathrm{AuM}_{it} + \upsilon_{it}$, by running two panel regressions of the respective forms

(14) $$\log\left(n_{it}\right) = a_i + b \cdot \log\left(\mathrm{AuM}_{it}\right) + \gamma\left(\widehat{n_{it}} - \exp\left\{\widehat{\log\left(n_{it}\right)}\right\}\right) + \varepsilon_{it},$$

(15) $$n_{it} = \alpha_i + \beta \cdot \mathrm{AuM}_{it} + \delta \cdot \left(\widehat{\log\left(n_{it}\right)} - \log\left(\widehat{n_{it}}\right)\right) + \upsilon_{it}.$$

Both equations add the difference in predictions (fitted values) with respect to the competing model to the model under the null hypothesis. If γ (resp. δ) is significantly different from zero, then the model based on log data (resp., untransformed data) is rejected. Running both panel regressions yields $\hat{\delta} = 11.01$ with $t\left(\hat{\delta}\right) = -4.36$ and $\hat{\gamma} = 0.00$ with $t\left(\hat{\gamma}\right) = 0.62$. In short, we can reject the model using untransformed data but cannot reject the model using log-transformed data. Thus the functional form proposed in our theoretical model is in line with the empirical findings.

Finally, we need to show that our log-linear model is indeed linear; hence we must test for threshold effects.[15] The idea here is that regression slopes might differ across regimes, where a regime is characterized by a (scalar) threshold value of a relevant "break" variable. What would justify the break points in our model? What circumstances would lead to these regimes? When assets under management grow, a FoF can increase the number of its target funds as frictional diversification costs become proportionally smaller. However, there may not exist enough target funds that satisfy this FoF investor's criteria. The FoF may be unconvinced of the target's ability to deliver high (risk adjusted) excess returns, or the FoF may face regulatory requirements that its current management is not trained to address. For these reasons and others, there is likely a limit to the number of funds actually eligible for inclusion in a FoF; that is, the number of viable additional holdings is *not* increasing in assets under management.

Following Hansen (1999), we estimate a threshold FE panel regression of the form

---

[15] A simple Ramsey reset test (including higher-order powers for the fitted values of the dependent variable and testing for their joint significance) does not indicate nonlinearity. In other words, the higher powers are neither individually nor collectively significant.

(16) $$\log\left(n_{it}\right) = a_i + b \cdot \log\left(\mathrm{AuM}_{it}\right) + \lambda \cdot \log\left(\mathrm{AuM}_{it}\right) \cdot D_{it} + v_{it}$$

for

(17) $$D_{it} = \begin{cases} 1 & \text{if } \mathrm{AuM}_{it} \geq \gamma, \\ 0 & \text{otherwise.} \end{cases}$$

Here γ is the threshold of AuM_{it} that activates an indicator variable D_{it}. To estimate the unknown break point, we perform a grid search. More precisely, we use 400 quintile values as candidates for γ while estimating equation (16) a total of 400 times. For each candidate value of γ we calculate an *F*-ratio of the form

(18) $$F(\gamma) = \frac{\hat{\varepsilon}'\hat{\varepsilon} - \hat{v}'\hat{v}}{\hat{v}'\hat{v}} \frac{1}{\#\mathrm{Obs.}},$$

here $\hat{\varepsilon}$ is the vector of residuals from (8), $\hat{v}$ is the vector of residuals from (16) for a particular γ, and $\#\mathrm{Obs.}$ is the number of valid observations. A structural break in the proposed log-linear relationship is suspected for $\hat{\gamma} = \arg\max\left(F(\gamma)\right) = 17.84$. But when is $F\left(\hat{\gamma}\right)$ statistically significant? We infer the critical value for (18) from the results of bootstrapping 5,000 times the residuals from our empirical estimate of (8), that is, under the null of no threshold effects. We thereby create 5,000 new data sets $\log(n_{it}), \log\left(\mathrm{AuM}_{it}\right)$ by adding the bootstrapped residuals to our base model. Within each of these data sets we again perform the grid search just described to maximize the value of equation (18), thus creating new estimates of $F_1, F_2, ..., F_{5000}$. This procedure yields a distribution for (18) from which we can now calculate the correct *p*-value for F. In our example, $\hat{\gamma} = 17.84$ with a test statistic of $F = 33.84$. After 5,000 resamplings we finally obtain the distribution of our test statistic under the null of no breaks. The *p*-value for F is then 0.103; hence we cannot reject the null hypothesis (of no break) at the 90% confidence level. Thus we can conclude that log-linear model is indeed linear suggesting that FoFs seem to be capable to hire (fire) individual funds that (do not) fulfill their selection criteria when their total assets under management increase (decline).

Figure 3

Influential FoFs

This figure shows estimates for b in $\log(n_{it}) = a_i + b \cdot \log(\text{AuM}_{it}) + \varepsilon_{it}$, where the ith FoF is dropped from the data set. Each data point reflects a regression beta estimated using 126 of 127 FoFs (i.e., with the ith FoF dropped). The gray line marks the full-sample estimate. Entries are sorted from left to right by the number of available fund observations. The top two printed lines give the number of quarterly observations available for a group of FoFs and the number of FoFs that belong to the indicated group.

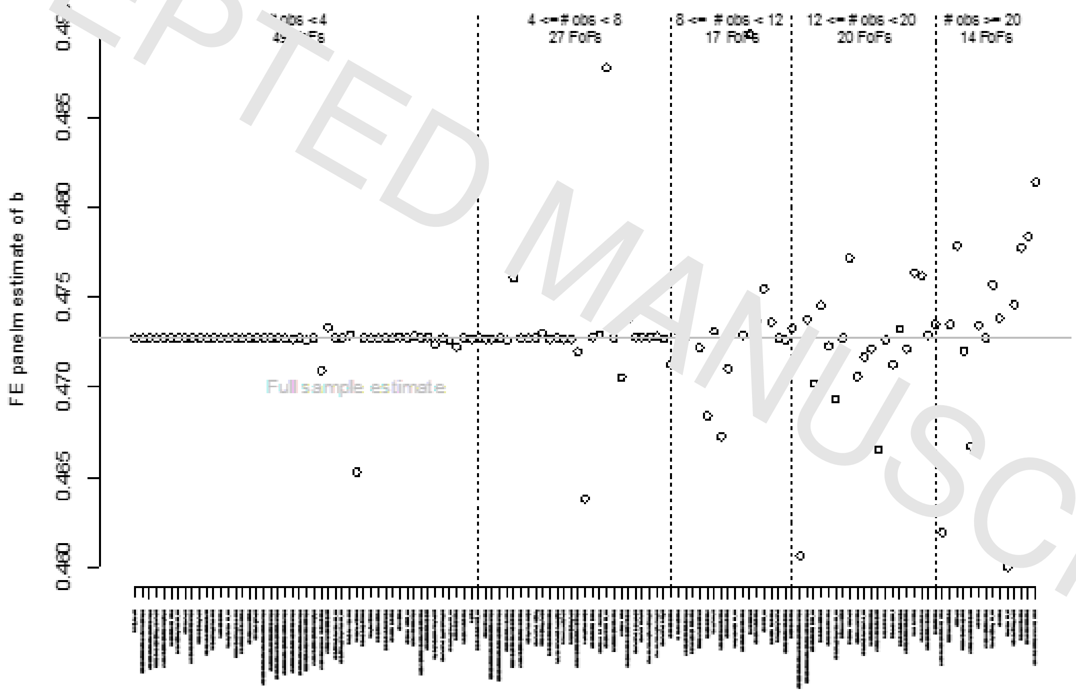


Figure 4

Influential Data Points

This figure plots the Cook's distance together with its 99% confidence level (indicated by the dashed gray line; values above this line represent possibly influential data points).

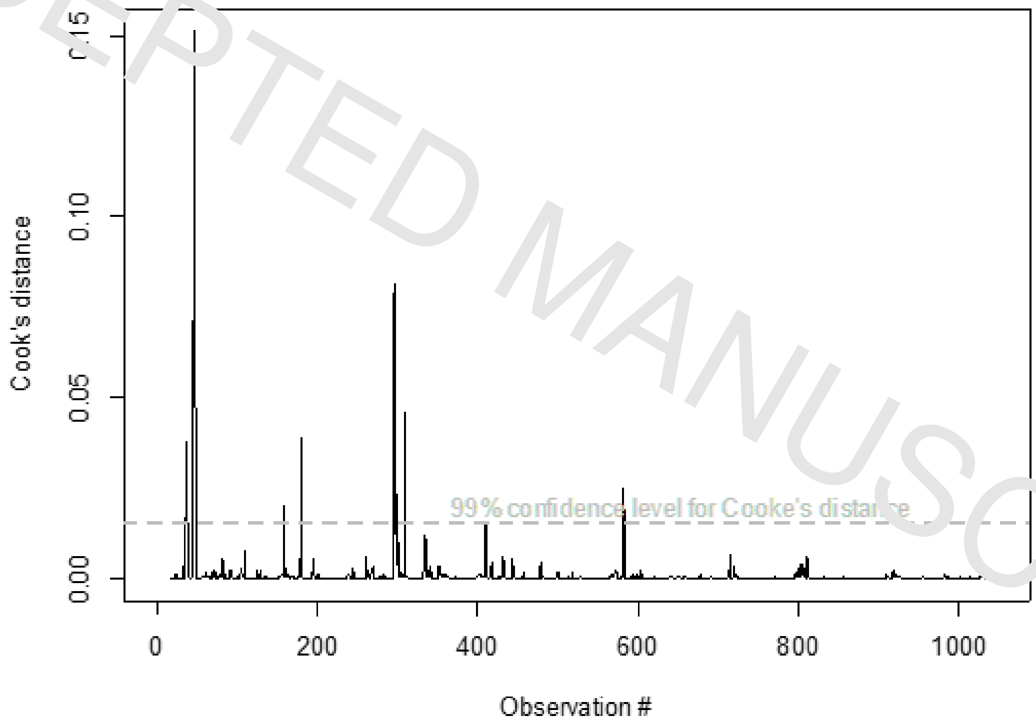
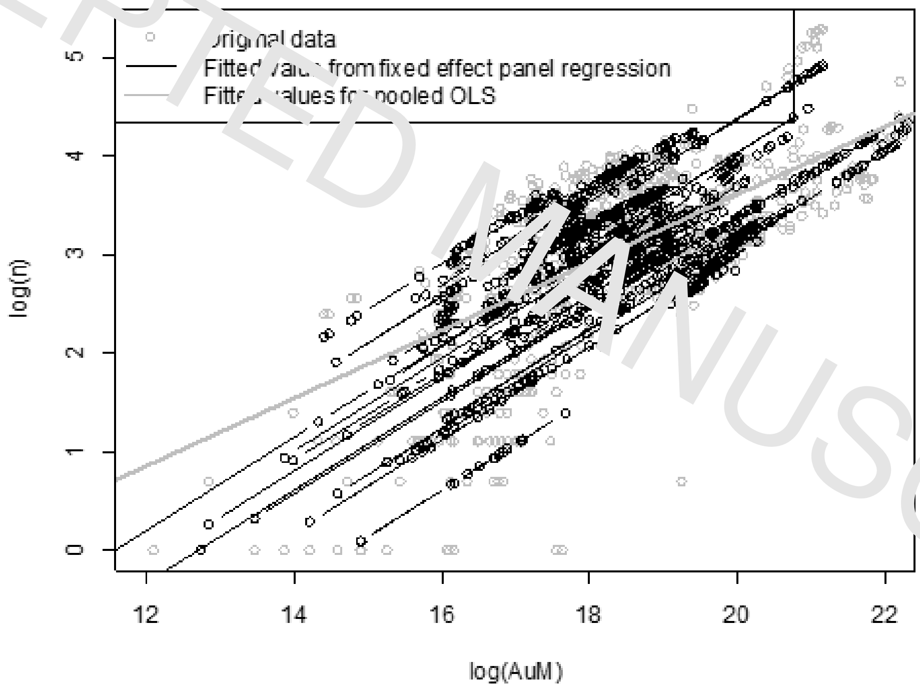


Figure 5

Number of FoF Holdings versus Assets under Management

Each circle represents one pair of observations on $\log(\mathrm{AuM})$ and $\log(n)$ for our FoF data set. Fitted values for the FE panel regression $\widehat{\log(n_{it})} = \hat{a}_i + \hat{b} \cdot \log(\mathrm{AuM}_{it})$ and the pooled OLS regression $\widehat{\log(n_{it})} = \hat{\alpha} + \hat{\beta} \cdot \log(\mathrm{AuM}_{it})$ are marked by solid black and gray lines, respectively.



5. Diversification and Performance

In this section, we investigate the relationship between the degree of diversification and FoF future performance as well as operational risk. In previous sections, we show that on average $\hat{b}$ is close to 0.5, i.e., FoFs on average behave as if they would employ a naïve diversification model as above. However, we also report that mean group (across buckets with similar number of observations) estimated betas show some variation, i.e., some funds deviate from the optimal level of diversification. Therefore, we test whether these violations of diversification (according to a naïve model) are related to deteriorating performance or more practically can be used to forecast future FoF returns or operational risk. We conjecture that over-diversification (too high costs, i.e., too low expected return per unit of risk or sloppy due diligence if costs are not spend) is as detrimental to performance as under-diversification (too high risk per unit of expected return). If underperformance (measured against various benchmark models) is related to over or under-diversification, investors can use this knowledge to better select FoF's. It also would offer additional support for (5). We propose for each individual FoF a diversification measure that classifies whether a FoF is (i) well diversified, (ii) underdiversified or (iii) overdiversified. The FoF is well diversified when $H_0 : b = 0.5$ is based on the two-sided test at a 10% significance level, while the FoF is underdiversified (overdiversified) when $\hat{b} < 1/2$ ($\hat{b} > 1/2$) based on the one-sided test at a 5% significance level. For every FoF, we estimate its diversification measure by using equation (8). Given that our time-series are relatively short, the use of the *t*-test instead of point estimates seems appropriate. The *t*-test statistic is a pivotal statistic with better sampling properties and, it should provide a correction for spurious outliers by normalizing the estimated parameter by the estimated variance of the parameter estimate.

We first employ both portfolio sorts and nonparametric cross-sectional regressions and then turn to multivariate panel regressions. Nonparametric regressions are particularly well suited for our purposes given that we aim to investigate whether there is a nonlinear relationship between the degree of diversification and the FoFs' performance, while a standard portfolio sort methodology allows us to gauge economic significance of our results. Given that portfolio sorts and nonparametric regression work well only for one variable (i.e., our diversification measure), we use multivariate regressions to control for the role of other variables that has been documented to explain FoF performance.

To ensure that our performance evaluation results are robust, we use various benchmark models. Motivated by Jagannathan, Malakhov and Novikov (2010), we use the equal-weighted FoF portfolio as a first benchmark portfolio. Such a benchmark portfolio measures how hedge funds perform relative to other funds, but do not say anything about risk-adjusted performance. Therefore, we use both the Carhart's (1997) four-factor model and the Fung and Hsieh (2004) seven-factor model. The Carhart's (1997) contains four risk factors: the excess returns on value-weighted mark index (MARKET), the size factor (SIZE), the value factor (HML) and the momentum factor (UMD).[16] The Fung and Hsieh (2004) model contains seven risk factors; the excess return of the S&P 500 index (SP), the return of the Russell 2000 index minus the return of the S&P 500 index (SIZE), the excess return of ten-year Treasuries (CGS10), the return of Moody's Baa-rated corporate bonds minus ten-year Treasuries (CREDSPR), and the excess returns of look-back straddles on bonds (PTFSBD), currencies (PTFSFX) and commodities (PTFSCOM).[17] Finally, since Brown, Gregoriou and Pascalau (2012) show that some FoFs are exposed to left-tail risk, we use two alternative specifications including; volatility (VOL) and jump (JUMP) factors proposed by Cremers, Halling and Weinbaum (2015) and the equity option factors (OTM_CALL and OTM_PUT) developed by Agarwal and Naik (2004).[18]

5.1. Nonparametric Regression and Diversification

[16] We download the Carhart factors from Kenneth French webpage.
[17] We thank David Hsieh for making trend-following factors available in his webpage.
[18] We use data obtained from OptionMetrics to replicate returns on these factors.

We start by running nonparametric cross-sectional regressions using a robust version of local regression proposed by Cleveland (1979) and further developed by Cleveland and Devlin (1988). In doing so, we fit a locally weighted polynomial regression model:

$$Performance_i = f\left(Diversification_i\right) + \xi_i \tag{19}$$

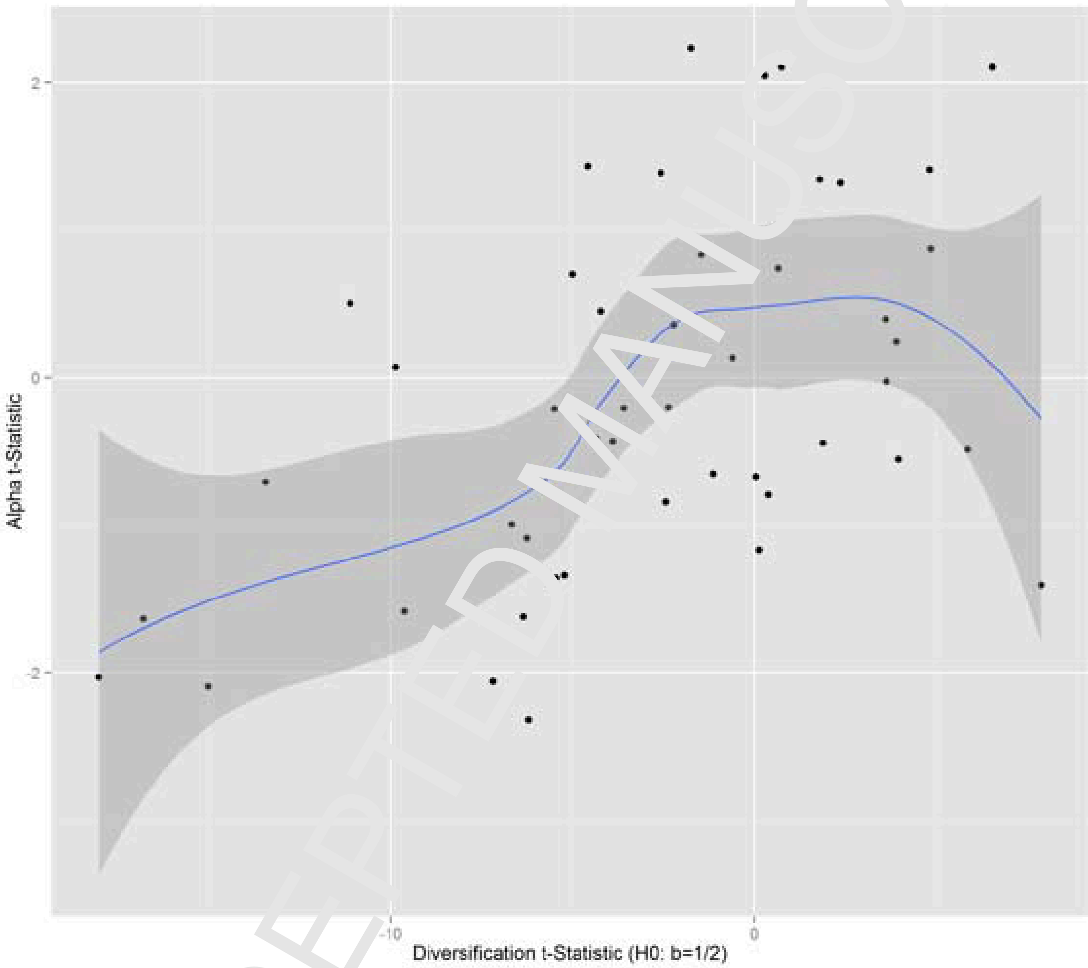
where $Performance_i$ is the *t*-statistic of alpha (intercept) obtained from the relative benchmark regression with respect to the equal-weight FoF index, the $Diversification_i$ is defined as the *t*-statistic of $\hat{b}$ estimated using equation (8) for every FoF *i* and $f()$ represents the locally weighted polynomial regression model.[19] At each point in the data set a low-degree polynomial is fitted to a subset of the data, with explanatory variable values near the point whose response is being estimated. The fitted values are computed by using the nearest neighbor routine and robust locally weighted regression of degree 1 with the tricube weight function.

Figure 6 presents fitted values for locally weighted polynomial regression results when the relative performance is regressed against the $Diversification_i$. From the scatterplot, we can observe that the FoFs diversifying along the lines of our simple model deliver better performance than the FoFs that hold only a few or a high number of individual hedge funds (relative to their AuM). The nonparametric regression also shows that the fitted value for alpha is highest where the $Diversification_i$ is close to zero, i.e., close to the null hypothesis. In summary, our results suggest that deviations from optimal diversification are penalized and there is – as conjectured – a nonlinear relationship between the diversification and relative performance.

[19] To obtain the *t*-statistic of alpha, we run for every FoF the following regression, $Ret_{i,t} = \alpha_i + \beta_i Ret_{EW,t} + \zeta_{i,t}$, where $Ret_{i,t}$ ($Ret_{EW,t}$) is the FoF's excess return (equal-weighted FoF portfolio return). We calculate the quarterly simple returns of individual hedge funds as in Aiken et al. (2013a). When a FoF holds the same hedge fund in two consecutive quarters with an unchanging cost, we calculate the underlying individual hedge fund's simple return as a fractional change in value. In case that several FoFs' holdings can be used to calculate such a return estimate, we use the median.

Figure 6
Diversification and Performance

This scatter plot presents locally weighted polynomial regression (LOESS) estimation results when each FoF's alpha *t*-statistics is regressed againts the FoF's diversification *t*-statisic. Alpha *t*-statistic is obtained regressing each FoF's returns againts equal-weighted FoF portfolio return. Diversification *t*-statistic is obtained by running a regression for each indivual fund and then testing whether *b=1/2*.



5.2. Portfolio Sorts and Diversification

To measure the potential economic value of our diversification measure, we use a standard portfolio sorting methodology. On each quarter from 2004Q3 to 2012Q2, we sort funds into three portfolios, based on the diversification measure,[20] and track the equally-weighted return produced by the portfolios in the following quarter. To assess the performance of the diversification measure sorted portfolios, we first calculate summary statistics of their quarterly raw returns. Such statistics have the benefit of being independent of the benchmark model. As straightforward statistics, we calculate the mean return, volatility and Sharpe ratio. For each mean return and Sharpe ratio, we test whether the difference between an equal-weight portfolio of optimally diversified FoFs and an equal-weight portfolio of FoFs that underdiversify (overdiversify) is statistically significant.[21]

Panel A of Table 2 reports that both mean return and Sharpe ratio are the highest for the optimally diversified portfolio, while the volatilities are quite similar across portfolios. Statistical tests show that mean spreads are significantly higher for optimally diversifying FoFs at the 1% level, while Sharpe ratio differences are significant at least at the 10% level. The relative performance results reported in Panel B support these findings. We find that both alphas and information ratios with respect to equal-weighted FoF portfolio are significantly higher for FoFs that diversify optimally. It also look like that the FoFs which underdiversify deliver the lowest performance among these three groups.

[20] Three portfolios that contain either (i) optimally diversified ($\hat{b} = 1/2$), (ii) underdiversified ($\hat{b} < 1/2$) or (iii) overdiversified FoFs ($\hat{b} > 1/2$).. FoF is optimally diversified when $\hat{b} = 1/2$ using two-sided test at a 10% significance level. The FoF is underdiversified (overdiversified) when $\hat{b} < 1/2$ ($\hat{b} > 1/2$) using one-sided test at a 5% significance level. $\hat{b}$ is estimated using an expanded window.

[21] We employ the Ledoit and Wolf (2008) approach to test difference between two Sharpe ratios.

We next turn to the performance of FoFs by using both the Carthart's (1997) four-factor model and Fung-Hsieh (2004) seven-factor model. Panel C shows that both the Carhart's alphas and information ratios are the highest for FoFs that diversify optimally. We do find that all for all three portfolios that FoFs are significantly exposed to value-weighted market portfolio, while other factors remains statistically insignificant. In addition, we can see that the magnitudes of risk loadings are quite similar across portfolios. As a complementary assessment of the performance of the sorted portfolios, we use the Fung and Hsieh (2004) model. Panel D shows that the alphas and information ratios are significantly higher for the portfolio containing the FoFs that diversify optimally. Again, the FoFs that underdiversify deliver the lowest performance. We do not find any significant difference between risk loadings among the three groups.

Given that Brown, Gregoriou and Pascalau (2012) show that FoFs are exposed to tail risk, we investigate issue by two additional specifications. First, using the both aggregate jump (JUMP) and volatility (VOL) risk factors proposed by Cremers, Halling and Weisbaum (2015). They show that stocks with high sensitivities to jump and volatility risk have low expected returns. Second, using the two option-writing factors developed by Agarwal and Naik (2004). Panel E and Panel D show that our results hold even after adding these factors. This suggests that our findings may not by driven by left-tail risk.

To investigate whether our portfolio sort results are consistent over time, Figure 2 plots the cumulative performance results for three diversification groups. Panel A plots cumulative excess returns, while in Panel B cumulative returns excess to equal-weighted FoF portfolio. Panel C and D plot the cumulative abnormal returns with respect to Carhart (1997) factors and Fung-Hsieh (2004) factors. Cumulative performance results are striking. These four panels show that FoFs that diversify along the simple predictions of model outperform the FoFs that either under- or overdiversify. Cumulative performance is consistently higher for them no matter whether we use raw returns, relative returns or risk-adjusted returns in terms of Carhart (1997) or Fung and Hsieh (2004) model.

Table 3
Portfolio Sorts and Diversification

This table present the out-of-sample results for three portfolios that contain either (i) optimally diversified ($\hat{b} = 1/2$), (ii) underdiversified ($\hat{b} < 1/2$) or (iii) overdiversified FoFs ($\hat{b} > 1/2$). The FoF is optimally diversified when $\hat{b} = 1/2$ using two-sided test at a 10% significance level. The FoF is underdiversified (overdiversified) when $\hat{b} < 1/2$ ($\hat{b} > 1/2$) using one-sided test at a 5% significance level. $\hat{b}$ is estimated using an expanded window. "Mean" is an annualized mean excess return for a particular portfolio. "Std" is an annualized standard deviation. "Sharpe" is an annualized Sharpe ratio. "Alpha" is an annualized intercept obtained from the respective benchmark model. "TE" is an annualized tracking error defined as a standard deviation of residual term. "IR" is an annualized information ratio defined as a respective benchmark model intercept divided by the standard deviation of residual term. The Ledoit and Wolf (2008) approach is used in Sharpe and IR difference tests. *T*-statistics of parameter estimates are presented in parenthesis.

EW Portfolios	Mean	Std	Sharpe	Alpha	TE	IR
Underdiversified	1.932	7.447	0.259	-1.249	1.011	-1.235
	(0.65)			(-3.48)		
Optimal	5.835	7.546	0.773	2.667	1.721	1.550
	(1.80)			(4.27)		
Overdiversified	3.250	6.635	0.490	0.473	1.597	0.296
	(1.15)			(0.73)		
Difference tests						
Optimal - Underdiversified	3.903		0.514	3.916		2.785
	(4.62)		(3.75)	(4.73)		(5.02)
Optimal - Overdiversified	2.586		0.284	2.194		1.254
	(2.76)		(1.72)	(2.40)		(2.07)

Panel C: Carhart Adjusted-Returns

EW Portfolios	Alpha	TE	IR	Market	SMB	HML	UMD
Underdiversified	0.545	3.782	0.144	0.393	-0.063	-0.182	0.046
	(0.40)			(5.35)	(-0.44)	(-1.81)	(1.17)
Optimal	4.294	3.709	1.158	0.370	0.067	-0.149	0.007
	(4.24)			(4.18)	(0.45)	(-1.54)	(0.17)
Overdiversified	1.964	3.478	0.565	0.319	0.037	-0.121	-0.001
	(1.45)			(4.72)	(0.30)	(-1.30)	(-0.01)
Difference tests							
Optimal - Underdiversified	3.749		1.014	-0.023	0.129	0.032	-0.039
	(5.61)		(3.27)	(-0.94)	(1.94)	(0.75)	(-2.44)
Optimal - Overdiversified	2.329		0.593	0.051	0.030	-0.028	0.008
	(2.47)		(1.68)	(1.59)	(0.41)	(-0.52)	(0.43)

Table 2
Continued

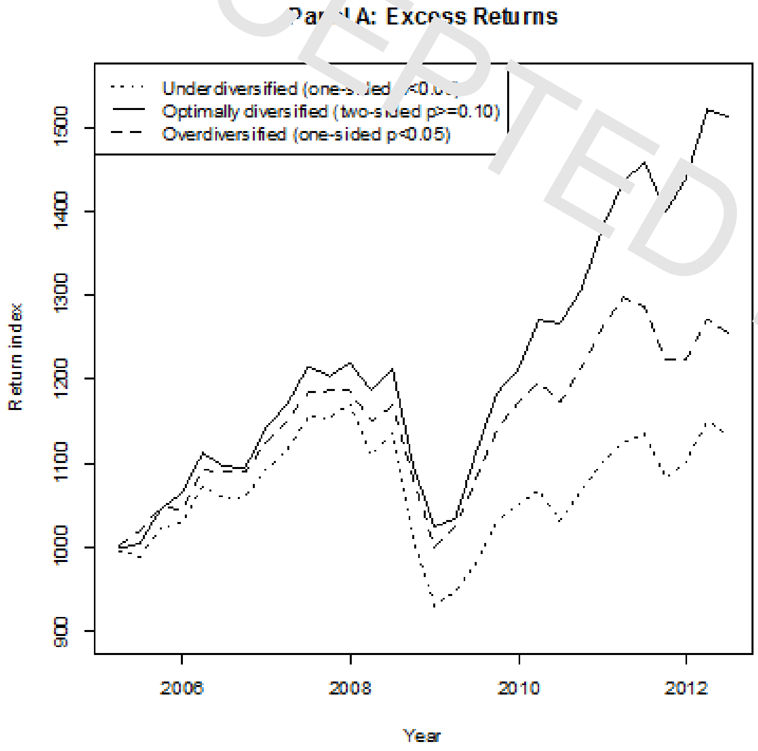
Panel D: Fung-Hsieh Adjusted-Returns

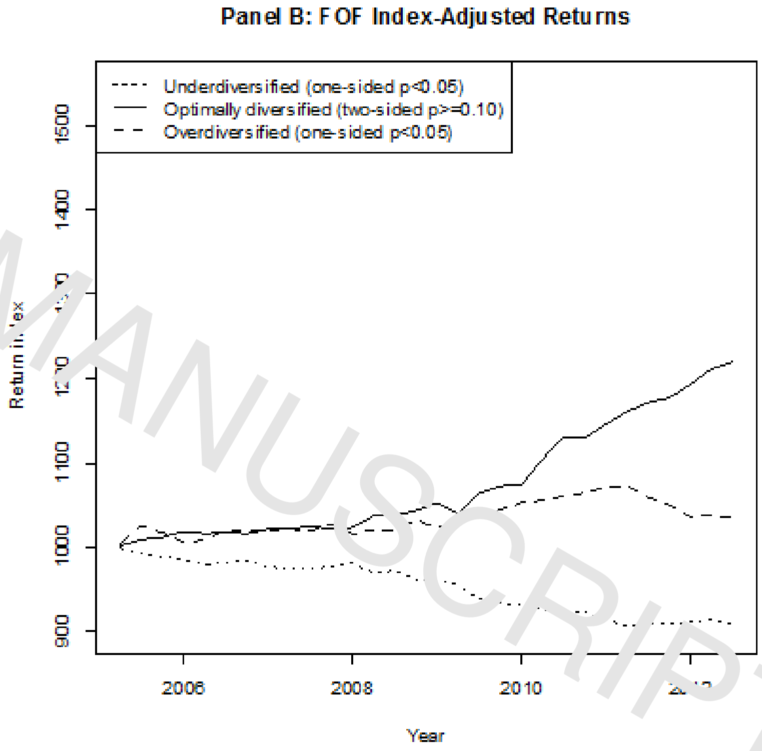
EW Portfolios	Alpha	TE	IR	SP	SCLC	CREDSPR	PTFSBD	PTFSFX	PTFSCOM	CGS10
Underdiversified	1.423	3.608	0.394	0.194	-0.179	0.040	-0.013	-0.022	-0.019	-0.096
	(0.92)			(2.86)	(-0.99)	(0.43)	(-0.64)	(-1.16)	(-0.96)	(-0.82)
Optimal	5.076	3.375	1.504	0.196	-0.022	0.129	-0.002	-0.006	-0.037	-0.059
	(3.18)			(3.77)	(-0.12)	(1.66)	(-0.10)	(-0.38)	(-1.82)	(-0.51)
Overdiversified	[illegible]	2.928	0.951	0.135	-0.060	0.125	-0.015	-0.002	-0.019	-0.127
	[illegible]			(1.98)	(-0.43)	(1.30)	(-0.89)	(-0.14)	(-1.06)	(-1.38)
Difference tests										
Optimal - Underdiversified	3.654		1.110	0.003	0.157	0.089	0.011	0.016	-0.018	0.037
	(4.62)		(3.13)	[illegible]	(3.09)	(2.03)	(1.07)	(1.75)	(-1.52)	(0.55)
Optimal - Overdiversified	2.292		[illegible]	[illegible]	0.038	0.004	0.013	-0.004	-0.017	0.068
	(2.49)		(1.59)	[illegible]	[illegible]	(0.11)	(1.05)	(-0.32)	(-0.97)	(0.79)

	Panel E: Cremers et al (2015) adjusted-returns						Panel F: Agarwal and Naik (2004) adjusted-returns.					
EW Portfolios	Alpha	TE	IR	Market	VOL	JUMP	Alpha	TE	IR	Market	Call_OTM	Put_OTM
Underdiversified	0.32	4.045	0.079	0.287	0.027	-0.048	[illegible]	4.47	0.113	0.327	-0.001	0
	(-0.18)			(-4.96)	(-0.98)	(-1.26)	[illegible]			(-4.54)	(-0.36)	(-0.19)
Optimal	3.937	3.61	1.091	0.293	-0.002	-0.043	[illegible]	[illegible]	[illegible]	0.34	0.001	0
	(-2.37)			(-6.85)	(-0.11)	(-1.24)	(-3.02)			(-6.30)	(-0.34)	(-0.15)
Overdiversified	1.753	3.52	0.498	0.267	0.000	-0.028	1.783	[illegible]	[illegible]	[illegible]	-0.001	-0.001
	(-1.15)			(-4.63)	(0.01)	(-0.95)	(1.22)			[illegible]	(-0.52)	(-1.16)
Difference tests												
Optimal - Underdiversified	3.616		1.011	0.006	-0.029	0.004	4.097		1.032	[illegible]	[illegible]	0.000
	(4.61)		(3.80)	(0.21)	(-1.82)	(0.66)	(5.28)		(3.14)	(0.26)	(0.60)	[illegible]
Optimal - Overdiversified	2.183		0.593	0.027	-0.003	-0.016	2.817		0.658	0.045	0.002	0.001
	(2.19)		(1.85	(0.98)	(-0.16)	(-1.49)	(3.63)		(1.97)	(1.32)	(0.89)	(1.83)

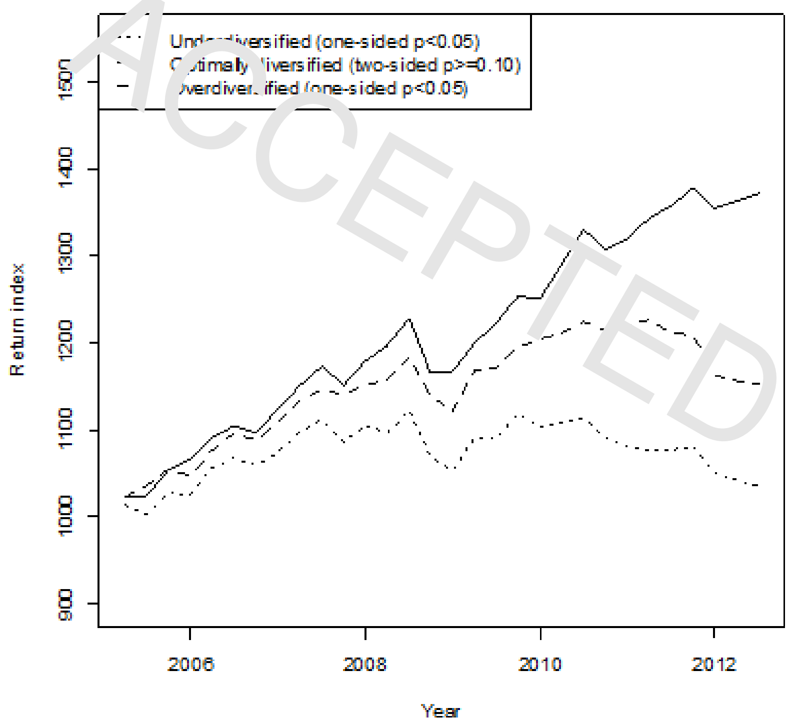
Figure 7
Out-of-sample performance and Diversification

These figures present the out-of-sample cumulative (Panel A) excess returns, (Panel B) FoF index-adjusted returns, (Panel C) Carhart-adjusted returns, and (Panel D) Fung-Hsieh-adjusted returns for three portfolios that contain (i) optimally diversified, (ii) underdiversified or (iii) overdiversified FoFs. FoF is optimally diversified when $\hat{b} = 1/2$ using two-sided test at a 10% significance level. The FoF is underdiversified (overdiversified) when $\hat{b} < 1/2$ ($\hat{b} > 1/2$) using one-sided test at a 5% significance level. $\hat{b}$ is estimated using an expanded window.

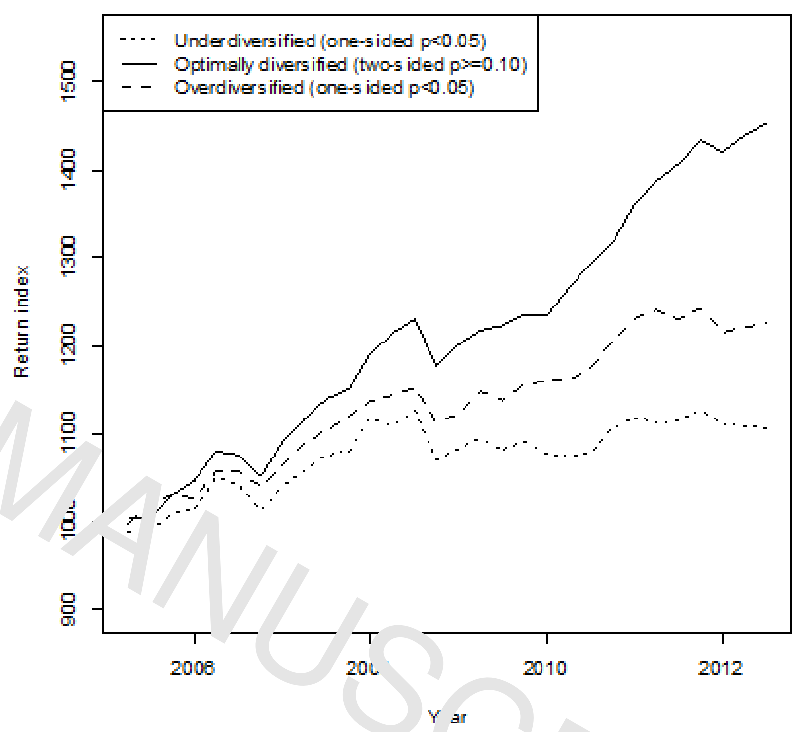




Panel C: Carhart-Adjusted Returns



Panel D: FH7-Adjusted Returns



5.3. Multivariate Analysis and Diversification

We next turn to the multivariate analysis of diversification and future FoF performance. For this purpose, we run a set of panel regressions to explore whether the variables documented by existing literature are more important variables in explaining the FoFs' performance than the proposed diversification measure. Given the quarterly frequency of our data, the most convenient way to examine the issue raised above is to run the following panel regression:

$$Return_{i,t} = \gamma_0 + \gamma_1 NonOptimal_{i,t-1} + \gamma_2 \log(AuM)_{i,t-1} + \gamma_3 \log(n)_{i,t-1} + \gamma_4 Problem_{i,t-1} + \gamma_5 BollenPool_{i,t-1}$$
$$+\gamma_6 DLR_{i,t-1} + \gamma_7 Flow_{i,t-1} + \gamma_8 Return_{i,t-1} + \gamma_9 StyleConcentration_{i,t-1} + \gamma_{10} TimeIn\text{var}iantControls_i,$$

where $Return_{i,t}$ is the FoF's quarterly relative return defined as the excess return over the equal-weighted FoF portfolio. To obtain the diversification measure on a quarterly basis, we create a dummy variable ($NonOptimal_{i,t-1}$) from our time series of diversification measures (estimated from OLS using an expanding data window). This dummy variable is defined as the statistically significant non-optimal diversification at the 10% level.[22] Thus, in this regression, the negative coefficient for $NonOptimal_{i,t-1}$ indicates a higher performance for FoFs that diversify along the lines of our simple model. We include quarterly dummies across specifications (not tabulated), and cluster standard errors by FoF.

Table 3 shows a robust and a negative relationship between $NonOptimal_{i,t-1}$ and FoF performance even once we add a set of control variables. It provides evidence that the FoFs that do not diversify along the simple model deliver 1.4%–3.4% lower relative returns per annum compared to the FoFs whose behavior is consistent with the model taking frictional costs into account.

To examine the robustness of the diversification and performance result, we base our control variables to the previous literature on both hedge fund performance and mutual fund performance.[23] First, we include the logarithm of the FoF's size ($\log(AuM)_{i,t-1}$) into the regression model. Brown, Frazier, and Liang (2008a) argue that only relatively large FoFs can absorb fixed costs that are required for setting up an effective due diligence process. Table 4 shows that the FoF's size does not drive our results. Although we

---

[22] Our results are quantitatively similar for the significance levels 1% and 5% as well as full-sample diversification measures. We also find quantitatively similar results using the Fama and MacBeth regressions (1973). These results are available upon a request.

[23] Table A2 provides the summary statistics for the set of control variables.

often find a significant and positive relation between $\log(AuM)_{i,t-1}$ and FoF relative performance, the coefficient for $NonOptimal_{i,t-1}$ remains negative and significant.

Second, we add $\log(n)_{i,t-1}$ into the regression model due to the fact that Pollet and Wilson (2008) document evidence suggesting that better diversification measured by a larger number of stocks is associated with superior performance especially for small-cap mutual funds. In contrast, we find some evidence that a lower number of underlying hedge funds is associated with greater relative performance. Importantly, adding $\log(n)_{i,t-1}$ does not impact the significance of our key variable $NonOptimal_{i,t-1}$.

We next include three different operational risk measures into our regression model.[24] The first one is defined as a fraction of "problem" funds that a FoF holds during the quarter ($problem_{i,t-1}$).The rationale stems from a series of papers done by Brown, Goetzmann, Liang, and Schwarz (2008b, 2009, 2012). Second, in the spirit of Bollen and Pool (2012), we form the suspicious return –flags variable that is defined as the fraction of underlying funds with suspicious return patterns. Finally, we build a variable that is defined as the fraction of underlying funds which have impose discretional liquidity restrictions (Aiken et al 2015b). Even after controlling for the role of three different aspects of operational risk, we find that the coefficients for $NonOptimal_{i,t-1}$ are negative and statistically significant.

Since Fung, Hsieh, Naik and Ramadorai (2008) document that the capital inflow attenuates the ability of good performing funds to deliver superior future performance, we add the FoF's quarterly capital flow ($flow_{i,t-1}$) into the diversification-performance regression model. After adding flow, we still find a negative and significant coefficient for $NonOptimal_{i,t-1}$ suggesting our finding is robust.

It is well-known that hedge fund monthly returns tends to be autocorrelated (e.g. Getmansky, Lo and Makarov 2004). In our case, the fund returns are quarterly, which according to Asness, Krail and Liew (2001) naturally alleviates the effects of non-synchronous price reactions on estimates of volatility and correlation. However, to ensure that the potential presence of stale prices does not contaminate our regression coefficients, we add, in the spirit of Dimson (1979) and Scholes and Williams (1997), the lagged return ($Return_{i,t-1}$) in our regression model. Consistent with idea of stale pricing, we find a

[24] Please, see from the Data Appendix details how operational risk variables are constructed.

positive and significant coefficient for lagged return. More importantly, the coefficient for $NonOptimal_{i,t-1}$ is still significantly negative.

Since the previous literature documents that the hedge fund compensation structure (e.g., Agarwal, Daniel and Naik 2009) and share restrictions (e.g., Aragon 2007) are associated with fund performance, we control for the role of these variables in our regression model. Our returns are net of underlying hedge fund fees, but gross of FoF fees. Hence, the compensation structure variables should control for the role of fee difference between FoFs. The FoF's with tighter share restrictions may manage capital flows better than the FoFs more honorable redemption restrictions. Even after adding a set of compensation and share restriction variables, we find that the coefficient for $NonOptimal_{i,t-1}$ is negative and statistically significant.

Our final concern is that some of the FoF may focus only on some of the investment styles, and thereby have concentrated portfolios. One of the shortcoming our diversification model is that it does not take this possibility explicitly into account, therefore we add the style concentration variable into our regression model. After controlling for the role of style concentration, our main result hold.

To summarize, our multivariate evidence suggests that there is a robust nonlinear relationship between the proposed diversification measure and the FoF performance. Indeed, the FoFs that set their diversification policy close to $n^2 \propto aum$. i.e., diversify more closely to our simple model tend to deliver relatively higher performance than the FoFs that hold either a few or a high number of individual hedge funds relative to the FoF's size.

5.4. Diversification and Operational Risk

Next, we analyse the relationship between the degree of diversification and operational risk. We start by examining whether the FoFs those do not diversify along simple model fail more often than the FoFs following closely the predictions of our model. In doing so, we run a set of Probit models, in which the FoF failure is explained the non-optimal diversification ($NonOptimal_{i,t-1}$), while controlling simultaneously the role of the FoF size, number of funds, style concentration and operational risk as well as compensation structure and share restrictions variables.

Table 4 reports that across specifications the coefficients of non-optimal diversification ($NonOptimal_{i,t-1}$) are consistently positive and statistically significant. Our results reveal that the degree of diversification is associated with failure probabilities. Indeed, those FoFs diversifying closely along the lines of our simple model exhibit a lower probability to fail.

It is interesting to note that none of three operational risk measures that we use as controls are capable to predict FoF failures, but the degree of diversification do so even after these three operational risk measures are added to the Probit model. Among the control variables, only the FoF size and style concentration are associated with FoF failure. However, even after controlling for the role of these variables the degree of diversification is robustly related to FoF failures.

To better understand the link between the diversification and operational risk, we run a set panel regressions, in which the operational risk are predicted by the degree of diversification. We start by examining the link between the fraction of underlying funds having imposed discretional liquidity restrictions and non-optimal diversification. Across specifications, we find positive and statistical significant coefficients for non-optimal diversification suggesting that those FoFs which diversify non-optimally are likely to invest in underlying funds that impose discretionary liquidity restrictions. For other two operational risk measures (operational problems and suspicious return patterns), we do not find a significant relationship between them and the degree of diversification. Hence, the non-optimally diversifying FoFs select in their portfolios more of the underlying funds that subsequently impose discretionary liquidity restrictions such as gates and side pockets, but degree of diversification is not associated with two other forms of operational risk.

6. Conclusion

We employ a model of naïve diversification with frictional diversification costs to motivate a positively sloped log-linear relation between a fund of fund's number of holdings and the value of its assets under management. When applied to a previously unavailable data set, our model produces evidence that FoFs diversify in line with this naive advice. In fact, we find a log-linear with a regression coefficient of close to 0.5. Diversification is no free lunch but only available to those who can afford it. No indications of nonlinearities or threshold effects were evident, and our results do not change materially when individual data points or individual FoFs are dropped from the sample. Finally we provided evidence that those FoFs following more closely our simple model's predictions are able to provide better future performance and fail less often.

Table 4
Multivariate Analysis and Diversification

This table reports results for the multivariate regressions, where the FoFs' relative returns are explained by the quarter lagged diversification measure and a set of one quarter lagged control variables. The variable "Relative return" is the FoF's quarterly relative return defined as the excess return over the equal-weighted FoF portfolio. The variable "Non-optimal diversification" is an indicator variable getting one if the FoF is underdiversified (overdiversified) when $\hat{b} < 1/2$ ($\hat{b} > 1/2$) using one-sided test at a 5% significance level, and otherwise zero. "Log(AuM)" and "Log(n)" are the logarithm of assets under management and number of funds. "Operational risk" is the fraction of problem funds that the FoF hold. Problem funds are classified using an indicator variable that takes on a value of one if the fund denotes any action brought by a regulator or the courts in Question 11 of its ADV filing, and zero otherwise. "Flow" is capital flows in percentiles. "Time FEs?" refers to calendar fixed effects. "Clustered SEs?" refers to the standard error that are clustered at FoF level.

Panel A: Baseline specifications

	Relative Return						
Non-optimal diversification	-0.0085	-0.0075	-0.0072	-0.0048	-0.0059	-0.0068	-0.0045
	(-2.72)	(-3.56)	(-3.28)	(-3.06)	(-3.99)	(-2.85)	(-3.34)
log(*AuM*)		0.00[illegible]2	0.0014	0.0008	0.0007	0.0016	0.0006
		(-1.54)	(-1.76)	(1.13)	(0.96)	(-1.84)	(1.01)
log(*n*)		-0.00[illegible]4	-0.0091	-0.0071	-0.0061	-0.0113	-0.0053
		(-3.40)	(-2.79)	(-2.05)	(-1.89)	(-2.73)	(-2.89)
Problem			0.0071				
			(-0.69)				
Bollen-Pool Score				0.0347			
				(2.28)			
DRL					-0.0132		
					(-2.13)		
Flow						0.0152	
						(-1.81)	
Lagged Return							0.3222
							(2.96)

Panel B: Compensation structure and share restrictions

	Relative return
Non-optimal diversification	-0.0034
	(-2.01)
log(aum)	0.0011
	(1.7[illegible])
log(n)	-0.0017
	(-0.6[illegible])
Management fee	0.3850
	(1.30)
Sales load fee	0.[illegible]241
	(1.47)
Incentive fee dummy	-0.0019
	(-0.98)
Lockup dummy	-0.0042
	(-2.42)
Redemption period	0.0024
	(0.58)
Notice period	0.0339
	(2.68)
Payout period	-0.0060
	(-0.27)
Style concentration	0.0012
	(0.23)

Table 5
Non-optimal diversification and FoF failure

This table present Probit analysis results, in which the fund failure is explained by lagged non-optimal diversification and set of control variables. “FoF failure” gets a value 1 when the FoF is terminated and 0 if the FoF continues to report with SEC. The other variables are as in Table 3.

		FoF failure	
Non-optimal diversification	0.0198	0.0204	0.0222
	(2.48)	(2.4[illegible])	(2.31)
log(aum)	-0.0161	-0.0158	-0.0108
	(-4.10)	(-3.7[illegible])	(-1.53)
log(n)	0.0057	-0.0009	-0.0526
	(0.62)	(-0.10)	(-2.41)
Problem		-0.0036	
		(-0.09)	
Bollen-Pool		-0.1728	
		(-1.02)	
DLR		0.0694	
		(1.34)	
Management fee			-1.8440
			(-0.71)
Sales load fee			-0.4124
			(-0.80)
Incentive fee dummy			0.0047
			(0.38)
Lockup dummy			0.0049
			(0.33)
Redemption period			-0.0039
			(-0.18)
Notice period			-0.1948
			(-1.59)
Payout period			0.3134
			(1.73)
Style concentration			-0.0931
			(-2.38)

Table 6
Operational risk and non-optimal diversification

This table presents the panel regression results, in which there operational risk proxies (*DLR*, *Problem* and *Bollen-Pool*) are explained by lagged *Non-optimal diversification* and a set of control variables. The variables are defined in Table 3.

	DLR		*Problem*		*Bollen-Pool*	
Non-optimal diversification	0.027	0.022	-0.002	0.000	-0.001	0.000
	(3.55)	(2.33)	(-0.28)	(-0.02)	(-0.89)	(-0.13)
log(aum)	-0.005	-0.004	-0.007	-0.001	0.001	0.001
	(-1.66)	(-1.10)	(-1.92)	(-0.33)	(1.07)	(1.35)
log(n)	0.014	0.000	0.001	-0.012	0.003	0.003
	(2.06)	(-0.03)	(0.13)	(-0.96)	(1.28)	(1.39)
Management fee		2.240		1.618		-0.343
		(1.53)		(0.99)		(-0.91)
Sales load fee		-0.203		0.167		0.005
		(-0.57)		(0.54)		(0.08)
Incentive fee dummy		0.010		0.000		-0.001
		(1.15)		(-0.04)		(-0.48)
Lockup dummy		0.027		0.008		-0.002
		(3.18)		(0.97)		(-1.28)
Redemption period		0.006		-0.037		-0.005
		[illegible]		(-1.74)		(-1.62)
Notice period		-0.161		-0.036		0.035
		(-2.67)		(-0.61)		(2.40)
Payout period		0.090		0.124		-0.078
		[illegible]		(0.66)		(-2.03)
Style concentration		-0.018		0.018		-0.008
		(-0.63)		(0.76)		(-1.70)
Lagged DLR	0.904	0.859				
	(32.30)	(25.32)				
Lagged Problem			0.870	0.881		
			(34.92)	(36.72)		
Lagged Bollen-Pool					0.848	0.808
					(29.65)	(24.33)

Appendix

A. Data Gathering Process

We start our data construction by downloading all SEC EDGAR master filing indices from 1993 to 2012. Each SEC filer is identified by its central index key (CIK). For each CIK appearing in the master indices, we download its most recently filed NSAR filing (either NSAR-A or NSAR-B). This yields 3,032 NSAR-A filings and 2,328 NSAR-B filings, for a total of 5,360 NSAR filings. Each CIK is classified as a closed-end fund if the answer to question 76 of its most recently filed NSAR form is zero. This yields a list of 4,521 CIKs of closed-end funds.

For each of these CIKs, we download all their annual reports (N-CSR, available from filing quarter 2003Q1), semi-annual reports (N-CSRS, available from filing quarter 2003Q3) and quarterly reports (N-Q, available from filing quarter 2004Q3). The intersection of these report types is available from filing quarter 2004Q3. These reports have a similar structure, so we use a single automated program to parse the fund holdings from each report. For each holding, we parse the name of the asset held, its original cost, and its current market value. To narrow the sample from closed-end funds down to funds of hedge funds, we restrict the sample to funds whose name indicates that the fund is 1) a fund of funds, 2) a fund of hedge funds, 3) a limited liability company (LLC), or 4) a limited partnership (LP). This leaves us with 202 CIKs. Finally, we manually inspect the holdings of these funds, and remove funds that do not primarily hold hedge funds. This leaves us with 127 distinct FoFs that have invested in 1,751 unique target funds. 776 of these funds we were able to match by name to a consolidated version of commercial hedge fund databases (BarclayHedge, EurekaHedge, HFR, Lipper TASS and Morningstar) created by following the steps described in Joenväärä, Kosowski and Tolonen (2015).

Next, we parse the underlying funds' investment strategies, discretionary liquidity restrictions and non-discretionary liquidity restrictions. Since it is very difficult to parse these variables using a single automated parser, we manually verify that we have gathered all of these variables correctly. This labor incentive step allows us also to check that the name of the asset held, its original cost, and its current market value are collected correctly without obvious parsing errors.

B. FoF-level Compensation Structure and Liquidity Variables

We use each FoF registration statement forms (N-2 and N-2 amendments) to gather information on compensation structure and share restriction variables. We hand collect variables related to fixed asset-based fees (*management fee*, *sales load fee*) as well as variables related to profit-based incentive fees (*incentive fee*, *hurdle rate*, and *high-water mark*). Information on used share equalization methods and crystallization periods is often missing or inaccurate. This implies that it would be difficult to estimate FoFs' gross returns. Among the share restriction variables, we obtain information on length of *lockup period*, *redemption period*, *notice period* and *payout period*. Table A1 presents summary statistic and Table A3 correlation matrix for N-2 variables. Since the incentive fee, high-water mark and hurdle rate are highly correlated with other, we never put them into same regression model. We opt to use the incentive fee as main control variable for managerial incentives. The conclusions are not sensitive to that choice.

C. Style Concentration and Operational Risk Measures

First, we construct a style concentration (Herfindahl-Hirschman) Index by gathering the underlying fund investment strategies from N-Q, N-CSRS, and N-CRS filings. We harmonize the strategies into eight styles: Directional, Event Driven, Long/Short, Market Neutral, Multi-Strategy, Relative Value, Sector and Other. Thereafter, we define the Herfindahl-Hirschman Index as the percentage of AuM weights of eight underlying fund strategies (*Directional*^{2} + *Event Driven*^{2} + *Long/Short*^{2} + *Market Neutral*^{2} + *Multi-Strategy*^{2} + *Relative Value*$^{2+}$ *Sector*^{2} + *Other*^{2}).

Next, we construct three measures for operational risk: discretionary liquidity restrictions, operational problems and suspicions return patterns. Table A1 and A2 presents summary statistics, whereas Table A2 presents the correlation matrix for these variables.

To find out which of the underlying hedge funds have imposed discretionary liquidity restriction (DLR), we follow Aiken et al. (2015b) classification. We define that the underlying fund has imposed a DLR when any FoF reports a position for the underlying fund that is 1) in a side pocket (either completely or partially), 2) subject to investor level gates, 3) liquidating, 4) organized as a special purpose vehicle or special liquidating vehicle, or 5) explicitly said to be illiquid or having its liquidity restricted. We define a FoF-level discretionary liquidity restriction (DLR) variable as the fraction of underlying funds which have impose discretional liquidity restrictions.

Next, we build the suspicions return patterns flags in the sprint of Bollen and Pool (2012). For each of the underlying fund that we can match to commercial database, we compute nine Bollen and Pool (2012) measures. Using a rolling 24-month return window, we compute; 1) the z-score of a return discontinuity around zero (in local currency); 2) the maximum adjusted R2 from a regression of excess fund returns against all possible subsets of the Fung and Hsieh (2004) risk factors; 3) the *p*-value of the slope from a regression of excess fund returns against the returns of a hedge fund's style index; 4) the *z*-score of a fund's return autocorrelation coefficient; 5) the *z*-score of a fund's return autocorrelation coefficient, conditioned on relative return level; 6) an indicator for whether a fund's history includes a string of at least three identical returns (in local currency, rounded at four decimals); 7) the χ^2 statistic against the null hypothesis of a uniform distribution for the last digit of returns (in local currency); 8) an indicator for whether the fund reports at least two exact zero returns (in local currency); 9) the percentage of negative returns (in local currency). We define a FoF-level Bollen-Pool score as the fraction of underlying funds which have suspicious return patterns.

Finally, following Brown, Goetzmann, Liang, and Schwarz (2008b, 2009), we classify as "problem" funds those individual hedge funds that answered "yes" to at least one question in Item 11 of ADV filing. Item 11 identifies all problems that the management or the related advisory affiliates have, including felonies, investment-related misdemeanors of any agency, SEC, CFTC, or self-regulatory issues, regulatory disciplinary action as well as civil lawsuits. We define a FoF-level problem as the fraction of underlying funds which exhibits problems.

Table A1
Cross-Sectional Descriptive Statistics

			N-2 variables			
Management fee	71	0.005	0.020	0.012	0.013	0.003
Sales load fee	64	0.000	0.050	0.015	0.020	0.014
Incentive fee	65	0.000	0.150	0.035	0.000	0.045
Incentive fee dummy	71	0.000	1.000	0.380	0.000	0.489
Hurdle dummy	71	0.000	1.000	0.169	0.000	0.377
High-water mark	71	0.000	1.000	0.366	0.000	0.485
Lockup dummy	59	0.000	1.000	0.508	1.000	0.504
Redemption period	59	0.250	1.000	0.364	0.250	0.169
Notice period	56	0.000	0.360	0.128	0.086	0.077
Payout period	59	0.000	0.164	0.078	0.082	0.040
			Operational risk (time-series mean)			
DLR	81	0.000	0.639	0.166	0.152	0.135
Problem (RT)	79	0.000	0.649	0.216	0.194	0.138
Bollen-Pool	80	-0.090	0.126	0.063	0.060	0.032

Table A2
Time-Series Descriptive Statistics

	N	Mean	Median	Std
		Performance statistics		
Fund excess return (%)	1043	0.899	1.68[illegible]	4.065
Capital flow (%)	1042	1.102	-0.145	20.605
		Diversification measures		
Diversification t-statistic	1043	-1.512	-1.0[illegible]	5.307
Diversification dummy	1043	0.573	1.000	0.495
		Portfolio statistics		
AUM (MUSD)	1283	282.315	94.377	681.813
log (AUM)	1283	18.3[illegible]	18.363	1.409
N	1283	28.4[illegible]5	25.000	19.927
log(N)	1283	3.156	3.219	0.655
Style concentration	1282	0.4[illegible]1	0.378	0.235
		Operational risk		
DLR	1[illegible]	0.175	0.100	0.199
Problem (RT)	1252	0.206	0.179	0.166
Bollen-Pool	1258	0.064	0.062	0.039

Table A3
Correlation Matrix

This table presents contemporaneous correlations from the 1,283 portfolios used in the out-of-sample analysis.

	Portfolio statistics				Operational risk			Fee structure (N-2)						Liquidity terms (N-2)			
	dx	log_aum	log_n	style_concentration	dlr	problem_rt	bollen_pool	management_fee	sales_load_fee	incentive_fee	incentive_fee_dummy	hurdle_dummy	highwater_mark	lockup_dummy	redemption_period	notice_period	payout_period
dx	1.00	0.02	0.10	-0.01	0.20	-0.09	0.03	-0.02	0.10	0.08	0.03	0.07	0.07	-0.05	0.02	0.14	0.12
log_aum	0.02	1.00	0.58	-0.09	0.03	-0.10	0.12	-0.17	0.14	0.09	0.06	0.05	0.11	-0.21	0.16	-0.31	0.17
log_n	0.10	0.58	1.00	-0.63	0.23	-0.03	0.19	0.04	0.13	0.10	-0.06	0.12	0.05	0.09	0.06	0.03	0.25
style_concentration	-0.01	-0.09	-0.63	1.00	-0.12	-0.03	-0.25	-0.14	-0.10	-0.10	0.09	-0.16	0.01	-0.23	-0.03	-0.18	-0.23
dlr	0.20	0.03	0.23	-0.12	1.00	0.04	-0.05	0.08	0.01	-0.11	-0.12	-0.11	-0.09	0.07	-0.02	-0.03	0.15
problem_rt	-0.09	-0.10	-0.03	-0.03	0.04	1.00	0.07	-0.04	-0.07	-0.04	-0.07	0.05	-0.05	0.00	-0.06	-0.06	-0.15
bollen_pool	0.03	0.12	0.19	-0.25	-0.05	0.07	1.00	0.06	0.13	-0.05	-0.14	0.09	-0.06	-0.05	-0.10	0.03	0.13
management_fee	-0.02	-0.17	0.04	-0.14	0.08	-0.04	0.06	1.00	0.05	-0.38	-0.36	-0.29	-0.33	-0.04	-0.24	0.00	0.10
sales_load_fee	0.10	0.14	0.13	-0.10	0.01	-0.07	0.13	0.05	1.00	0.42	0.38	0.20	0.43	-0.05	0.19	0.09	-0.02
incentive_fee	0.08	0.09	0.10	-0.10	-0.11	-0.04	-0.05	-0.38	0.42	1.00	0.91	0.64	0.90	-0.05	0.31	0.24	-0.08
incentive_fee_dummy	0.03	0.06	-0.06	0.09	-0.12	-0.07	-0.14	-0.36	0.38	0.91	1.00	0.53	0.96	-0.09	0.36	0.13	-0.20
hurdle_dummy	0.07	0.05	0.12	-0.16	-0.11	0.05	0.09	-0.29	0.20	0.64	0.53	1.00	0.45	-0.08	0.26	0.15	0.06
highwater_mark	0.07	0.11	0.05	0.01	-0.09	-0.05	-0.06	-0.33	0.43	0.90	0.96	0.45	1.00	-0.09	0.36	0.13	-0.20
lockup_dummy	-0.05	-0.21	0.09	-0.23	0.07	0.00	-0.05	-0.04	-0.05	-0.05	-0.09	-0.08	-0.09	1.00	-0.27	0.46	0.07
redemption_period	0.02	0.16	0.06	-0.03	-0.02	-0.06	-0.10	-0.24	0.19	0.31	0.36	0.26	0.36	-0.27	1.00	-0.17	-0.17
notice_period	0.14	-0.31	0.03	-0.18	-0.03	-0.06	0.03	0.00	0.09	0.24	0.13	0.15	0.13	0.46	-0.17	1.00	0.46
payout_period	0.12	0.17	0.25	-0.23	0.15	-0.15	0.13	0.10	-0.02	-0.08	-0.20	0.06	-0.20	0.07	-0.17	0.46	1.00

References

Agarwal, V., Daniel, N.D. and Naik, N.Y. (2009). Role of Managerial Incentives and Discretion in Hedge Fund Performance. The Journal of Finance, v64(5), pp. 2221–2256.

Agarwal, V., Fos, V. and W. Jiang (2013), Inferring Reporting-Related Biases in Hedge Fund databases from Hedge Fund Equity Holdings, Management Science, v59(6), pp. 1271–1289.

Agarwal V., Nanda, V., and Ray S., (2013). Institutional Investment and Intermediation in the Hedge Fund Industry. Georgia State University.

Agarwal, V. and Naik, N.Y. (2004). Risks and Portfolio Decisions Involving Hedge Funds. Review of Financial Studies, v17(1), pp. 63–98.

Agarwal V., Lu X., and Ray S., (2016). Under One Roof: A Study of Simultaneously Managed Hedge Funds and Funds of Hedge Hunds. Management Science, v62(3), pp. 722–740.

Aiken A., Clifford C.P. and Ellis J.A. (2013), Out of the Dark: Hedge Fund Reporting Biases and Commercial Databases. Review of Financial Studies, v26(1), pp. 208–243.

Aiken A., Clifford C.P. and Ellis J.A. (2015a), The Value of Funds of Hedge Funds: Evidence from Their Holdings. Management Science, v61(10), pp. 2415–2429.

Aiken A., Clifford C.P. and Ellis J.A. (2015b). Hedge Funds and Discretionary Liquidity Restrictions. Journal of Financial Economics, v116(1), pp. 197–218.

Amin G. and H. Kat (2002), Portfolios of Hedge Funds: What Investors Really Invest In, Working paper, ISMA University of Reading.

Aragon, G. (2007). Share Restrictions and Asset Pricing: Evidence from the Hedge Fund Industry. Journal of Financial Economics, v83, pp. 33–58.

Bollen N.P.B., Joenväärä J. and Kauppila M. (2018), Hedge Fund Performance Prediction. Working paper.

Bollen, N.P.B., and V. Pool (2012). Suspicious Patterns in Hedge Fund Returns and the Risk of Fraud. Review of Financial Studies v25, pp. 2673–2702.

Brennan M. (1975), The Optimal Number of Securities in a Risky Portfolio When There Are Fixed Costs of Transacting: Theory and Some Empirical Results, Journal of Financial and Quantitative Analysis, v10(4), pp. 483–496.

Brown S., T. Fraser, and B. Liang (2008a), Hedge Fund Due Diligence: A Source of Alpha in a Hedge Fund Portfolio Strategy, Journal of Investment Management, v6(2), pp. 22–33.

Brown S., W. Goetzmann, and R. Ibbotson (1999), Offshore Hedge Funds: Survival and Performance, 1989–95. Journal of Business, v72(1), pp. 91–117.

Brown S., Goetzmann W. and J. Park (2001), The Journal of Finance, v56(5), pp. 1869–1886.

Brown S., W. Goetzmann, B. Liang, and C. Schwarz (2008b), Mandatory Disclosure and Operational Risk: Evidence from Hedge Fund Registration, Journal of Finance, v63(6), pp. 2785–2815.

Brown S., W. Goetzmann, B. Liang, and C. Schwarz (2009), Estimating Operational Risk for Hedge Funds: The Ω-Score, Financial Analysts Journal, v65(1), pp. 43–53.

Brown S., W. Goetzmann, B. Liang, and C. Schwarz (2012), Trust and Delegation, Journal of Financial Economics, v103(1), pp. 221–234.

Brown S.J., G.N. Gregoriou and R. Pascalau (2012), Diversification in Funds of Hedge Funds: Is It Possible to Overdiversify? Review of Asset Pricing Studies, 2 (1), pp. 89–110.

Carhart, M., (1997), On Persistence in Mutual Fund Performance, Journal of Finance, 52(1), p. 57–82.

Chopra V. and W.T. Ziemba. (1993). The Effects of Errors in Means, Variances, and Covariances on Optimal Portfolio Choice. *Journal of Portfolio Management*, pp. 6-11.

Cleveland, W. S. (1979). Robust Locally Weighted Regression and Smoothing Scatterplots. Journal of the American Statistical Association v74(368), pp. 829–836.

Cleveland, W. S. and Devlin, S. J. (1988). Locally-Weighted Regression: An Approach to Regression Analysis by Local Fitting. Journal of the American Statistical Association v83(403), pp. 596–610.

Cremers, M., Halling, M. and Weinbaum, D. (2015). Aggregate Jump and Volatility Risk in the Cross-Section of Stock Returns. Journal of Finance, v70(2), pp. 577–614.

Edelmann D, Fung W and Hsieh D (2013), Exploring Uncharted Territories of the Hedge Fund Industry: Empirical Characteristics of Mega Hedge Fund Firms. Journal of Financial Economics v109(3), pp. 734–758.

Elton E.J. and M.J. Gruber (1977), Risk Reduction and Portfolio Size: An Analytic Solution, Journal of Business, v50(2), pp. 415–437.

Elton E.J., Gruber M.J. and J. Busse (2004), Are Investors Rational? Choices Among Index Funds, Journal of Finance, v59(1), pp. 261-287.

Fama, E. F. and MacBeth, J. D., (1973), Risk, Return, and Equilibrium: Empirical Tests, Journal of Political Economy, v81(3), pp. 607–636.

Fung W and Hsieh D. (1997), Survivorship Bias and Investment Style in the Returns of CTAs: The Information Content of Performance Track Records, Journal of Portfolio Management,v24, pp. 30–41.

Fung, W. and Hsieh D. (2000), Performance Characteristics of Hedge Funds and CTA Funds: Natural versus Spurious Biases. Journal of Financial and Quantitative Analysis 35, 291–307.

Fung, W. and Hsieh, D., (2004). Hedge Fund Benchmarks: A Risk Based Approach. Financial Analyst Journal v60, pp. 65–80.

Fung, W., Hsieh, D., Naik, N., Ramadorai, T., (2008). Hedge Funds: Performance, Risk, and Capital Formation. Journal of Finance, v63(4), pp. 1777–1803.

Frazzini A. and L.H. Pedersen (2014), Betting Against Beta, Journal of Financial Economics, v111(1), pp. 1–25.

Getmansky, M., Lo, A. and Makarov, I. (2004). An Econometric Model of Serial Correlation and Illiquidity in Hedge Fund Returns. Journal of Financial Econometrics, v74, pp. 529–609.

Goetzmann W. and A. Kumar (2008), Equity Portfolio Diversification, Review of Finance, v12(2), pp. 433–466.

Goldsmith D. (1976), Transaction Costs and the Theory of Portfolio Selection, Journal of Finance, v31(4), pp. 1127–1139.

Greene W. (2008), Econometric Analysis, 6th edition, Prentice Hall.

Greenwich Associates (2011), Doing the Due Diligence Detective Work, Investhedge.

Hansen B. (1999), Threshold Effects in Non Dynamic Panels: Estimation, Testing and Inference, Journal of Econometrics, v93(2), pp. 345–368.

Henker T. and G. Martin (1998), Naïve and Optimal Diversification for Managed Futures, The Journal of Alternative Investments, v1(2), pp. 25–39.

Huber P.J. (1981) Robust Statistics. Wiley.

Jagannathan R., Malakhov A., and D. Novikov (2010), Do Hot Hands Exist Among Hedge Fund Managers? An Empirical Avaluation. The Journal of Finance, v65(1), pp. 217–255.

Joenväärä, J., Kosowski, R., and P. Tolonen (2015), Hedge Fund Performance: What Do We Know? Imperial College Business School and University of Oulu.

Joenväärä, J., Kosowski, R., and P. Tolonen (2018), The Effect of Investment Constraints on Hedge Fund Investors Returns. Journal of Financial and Quantitative Analysis, Forthcoming.

Ledoit, O., and Wolf, M., (2008), Robust performance hypothesis testing with the Sharpe ratio, Journal of Empirical Finance, v15(5), pp. 850–859.

Lhabitant F.S. and M. Learned (2002), Hedge Fund Diversification: How Much is Enough? The Journal of Alternative Investments, v5(3), pp. 23–49.

Liang, B. (2000), Hedge funds: The Living and the Dead. Journal of Financial and Quantitative Analysis 35 (3), 309–326.

Mackinnon J, White H. and Davidson R. (1983), Tests for Model Specification in the Presence of Alternative Hypothesis: Some Further Results, Journal of Econometrics, v21(1), pp. 53–70.

Mitton T. and K. Vorkink (2007), Equilibrium Under-diversification and the Preference for Skewness, Review of Financial Studies, v20(4), pp. 1255–1288.

Panageas S. and M. Westerfield (2009), High-Water Marks: High Risk Appetites? Convex Compensation, Long Horizons, and PortfolioChoice, Journal of Finance, v62(1), pp. 1–36.

Pollet J. M. and Wilson M. (2008), How Does Size Affect Mutual Fund Behavior? Journal of Finance, v63(6), pp. 2941–2969.

Phillips P., Cathart A and Teale J. (2007), The Diversification and Performance of Self Managed Superannuation Funds, Australian Economic Review, v40(4), pp. 339–352.

Plokovnichenko V. (2005), Household Portfolio Diversification: A Case for Rank Dependent Preferences, Review of Financial Studies, v18(5), pp. 1467–1502.

Samuelson P. (1967), General Proof that Diversification Pays, Journal of Financial and Quantitative Analysis, v2(1), pp. 1–13.

Scherer B. (2013), Frictional Costs of Diversification, Journal of Portfolio Management, v39(3), pp. 7–9.

Statman M. (2004), The Diversification Puzzle, Financial Analysts Journal, v60(4), pp. 44–52.

Xu X. and Scherer, B. (2007), Performance based Fees and Risk Shifting with knockout barrier, Journal of Investment Management, v5(3), p. 1–18.